

Beth Kilss and Fritz Scheuren, Internal Revenue Service

One of the main reasons for the rapid growth in the statistical uses of administrative records during the 1970's was that certain data could be obtained at a relatively low cost and without increasing respondent burden [1]. Cost and respondent burden are likely to continue to be important factors in the statistical use of administrative records in the current decade given the recent changes in the Federal government's statistical programs [2]. There are very few examples of administrative record systems that have been designed with statistical uses in mind. Instead, statisticians, economists and researchers have had to locate existing administrative data that best suit their intended use. In other words, most statistical uses have been developed in an ad hoc manner. The statistician is not in control of the design and collection of the records so as to yield desired statistical characteristics. Consequently there are many unanswered questions about the quality of Federal statistics from administrative records.

This paper discusses the quality of information obtained from one of the Federal statistical data systems derived from administrative records--the Internal Revenue Service's Statistics of Income (SOI) program for individual income tax returns. The main theme is that quality concerns are growing--paradoxically at the same time as reliance on administrative records for statistical purposes has been increasing.

The material in the paper is divided into three main parts. The first of these is a brief analysis of the SOI program. This case study illustrates and quantifies a number of dimensions of what is meant by quality. It also makes concrete the usual processing and reporting environment in which statistics from administrative records are developed.

In the second part of the paper we discuss the impact on quality of dollar budget and burden budget cuts. Quite obviously, with the recent cuts in the statistical budgets, virtually all the statistical agencies and non-statistical agencies are threatened with a serious "attack" on the quality of statistics, whether from surveys or administrative records. The key for each agency is, of course, to deal with these cuts so as to minimize, to the greatest extent possible, their impact on final delivered quality.

The paper concludes with some observations about the kinds of issues that warrant consideration if the statistical community is to preserve the needed quality of Federal statistics given the likelihood of still further dollar budget and burden budget cuts.

1. STATISTICS OF INCOME (SOI) PROGRAM

Shortly after the passage of the constitutional amendment calling for the reinstatement of the income tax, IRS began tabulating infor-

mation from tax returns in its Statistics of Income program. (There was a Federal income tax during and shortly after the Civil War, which was declared unconstitutional. The income tax was reinstated in 1913 with the ratification of the Sixteenth Amendment to the Constitution and the enactment of the Revenue Act of 1913.)

One part of the Statistics of Income program involves collecting and processing data from the individual income tax returns. The program for these data has developed over the years and, now, microdata files for public release are produced annually by the SOI Division at the Internal Revenue Service. Statistical information is also tabulated and published in a volume of the Statistics of Income series [3]. The statistics produced in the SOI program for individuals are based on a stratified probability sample of unaudited individual income tax returns and represent coverage of the Forms 1040 and 1040A filed by U.S. citizens and residents for a particular income year. For Income Year 1979, the SOI sample consisted of about 200,000 returns taken from the 92.7 million that were filed [3].

Responsibility for the development of the tax forms for the coming filing season lies with the Tax Forms Committee which coordinates the activity with the Joint Committee on Taxation, with other interested members of the Treasury Department, as well as with the United States Congress. Internal Revenue Service provisions needed because of tax law changes and attempts at simplification are also coordinated with the Tax Forms Committee. In addition, there is now an involvement with the Office of Management and Budget (OMB) in this process as a result of the Paperwork Reduction Act, which requires that OMB clear all tax forms.

The SOI Division plays a minor role in the development of the tax forms. Usually, information about the usage of prior tax forms is provided by the Division to the Tax Forms Committee. During the Summer and Fall while the forms are getting into shape, the SOI Division is developing the statistical system which will process these forms. There is extensive consultation with SOI users, notably the Office of Tax Analysis, learning what information the users are expecting to obtain from the returns. Occasionally, the Division will request tax form changes to simplify its work or to meet user requirements. In any event, it cannot finalize its programs until the final forms have been decided.

Development of the computer systems for statistical purposes

The development of the statistical system that will process the tax year's forms, given the particular set of instructions for that

year, includes the development of sample design criteria, computer software, processing instructions, clerical training materials, and so forth. If there are major tax law changes affecting individual income returns then the Statistics of Income program undergoes a substantial redesign. (For a brief summary of major tax law changes affecting individual income returns from 1917 to 1979, see [4]. Tax law changes have an obvious detrimental impact on SOI in that some element of stability in a statistical system is a key to quality. It is hard for an agency which is employing statistics from changeable administrative records to achieve stability. To do that requires a different kind of ingenuity than at least we have been able to fully implement so far. The newer computer technologies which allow for generalized systems and the degree to which we can apply them will be a test of to what extent we can actually address these underlying systemic quality issues.

It might be noted in passing that surveys have a distinct advantage over administrative records regarding the development of computer systems for statistical purposes. In a survey, like the Current Population Survey (CPS) for example, the control of the questionnaire is more or less completely with the statistical agencies that are employing the form. In the case of SOI, that is not true, which increases the costs considerably.

The administrative data capture process and quality improvements

The administrative data capture and the supplementary SOI statistical data capture processes work hand in hand. Before describing the statistical part of this system in more detail, the nature of the quality checks that are imposed on the administrative system will be discussed briefly.

When returns are filed initially with the ten IRS service centers, they are processed for administrative purposes to determine the correct tax liability. As part of the key-entry process, there are a whole set of procedures known as "consistency tests" which are employed to identify mathematical errors made by taxpayers and mistakes made in the actual data capture process itself. (Consistency tests are tests as simple as adding up all income to see if it adds up to the total, calculating or recalculating the tax based on information from the return, and looking for unusually large or small reports for particular items.) In terms of mathematical errors made by taxpayers, two out of three of the over six million detected mathematical errors for 1979 were in favor of the taxpayer. This tendency for mathematical errors to favor the taxpayer has been such a longstanding historical phenomenon that perhaps we should just make an observation about it.

The first thing to dispel is the notion that the persistent direction of the errors made each year indicates that there is a wholesale attempt to bilk the government. On the contrary, most people, as a patriotic duty, try to pay their fair share; what may be happening is that there is a sense of frustration that

comes into play when we calculate our tax (as we sit there in the "dark of night," typically pictured in the cartoons, with a candle lit and a lot of scrap paper and pencils--several broken ones, perhaps--anyway, the erasers completely used up). During these calculations, we find an answer which is "reasonably small" and we think it must be right, and so we end up with that figure. If the number seems too high to us, we recalculate it again. And so there is, in a sense, a stopping-rule problem here. Taxpayers tend to stop when they get the answer they want, which is not always the right answer. And it is not a conscious attempt to cheat; it is simply part of the process of reporting on one's income.

Taxpayers do, as the conventional wisdom asserts, tend to underpay their taxes in general, rather than overpay them. There is considerable evidence suggesting that in fact there is a systematic underreporting of some income types and, consequently, underpayment of taxes. (See for example [5].) Developing a strategy for making this kind of information available to users of Statistics of Income data so that it is relevant to their use of SOI statistics is a very hard problem and one that has not been given enough attention for us to report on it here [6].

The elimination of mathematical errors, needed to ensure the accuracy of the information for administrative purposes, is also a quality improvement step in the SOI program. However, not all information collected on the individual income tax return is of equal importance for tax administration purposes. Resources needed to ensure accurate reporting and processing of information will be concentrated where administrative priorities lie. Therefore, reporting and processing errors can have a negative impact on the accuracy of items which may be very important for statistical applications, but not so for program administration.

Separate statistical operations

This section will briefly discuss the statistical operations which are separate from the administrative processes. Returns received in the service centers for administrative processing are examined for completeness and passed on to be transcribed and key-entered onto a disk and then onto tapes, called "transaction tapes," containing virtually all of the information on the main part of the return, and much of what is on the schedules that have to be attached. These transaction tapes are sent to the National Computer Center in Martinsburg, West Virginia, where they are, as the phrase is, "bumped up" against the Individual Master file (IMF) of all taxpayers. During this processing, certain information is posted to the IMF from the transaction tapes and comparisons are made with what taxpayers did in former years.

It is at this point that the sampling of the tax return information is done for statistical purposes, generating another computer tape to pull the returns which have been filed in the service centers after the

initial administrative data capture.

After the sampling at the National Computer Center, the selected records are sent back to the service centers on tape and the pulling of the actual sample returns begins. Additional errors can occur here, of course. For example, it may not be possible to find a particular return. In a large "paperwork factory" like a service center, returns are not always "locatable" (usually about 0.5% of the returns are unavailable for statistical processing). For one thing, they may have gone on to some other administrative function. To the extent possible, the return is copied or otherwise retained once it is selected, even though it may be needed almost immediately for an examination, or for some other taxpayer contact (for a continuing investigation, for example, in a fraud case).

For the bulk of the sampled returns, additional information is manually edited and transcribed for merging to the computer record obtained from the administrative processing. An example of an additional set of items might be the detailed calculation of the residential energy credit. Another illustration might be concerning the All Savers Certificate interest information, details of which might not all have been picked up, or not have been felt needed for administrative purposes, but are felt needed for tax policy examination of the new legislation.

Statistical testing in the service center is accomplished while the return is still available [7]. These tests are an important factor in the SOI program because certain kinds of errors are not uncovered in administrative processing as they do not impact on the administrative purpose. For example, a very low income taxpayer who may not have even needed to file a return might have a record with errors on it which did not get corrected because there were no tax consequences. These errors would impact on the statistics from the records, however, and so they need to be examined and an adjustment procedure developed for them.

When errors are found, the return is re-examined as needed, and an attempt is made to resolve the errors in the service centers. Some of the errors that are uncovered are taxpayer errors that are not necessarily resolvable using the return. In these cases, a judgment is needed that is based on the end use of the data, or at least the perceived end use. In order to control these judgments, and to carefully monitor how they are made, there is a later processing stage when all of the data on tape are shipped to the Detroit Data Center and combined and retested. (The Detroit Data Center is where all nontax processing is done at IRS such as statistical work after the initial data capture, payroll recordkeeping, etc.) Adjustments are made with the direct involvement of the subject-matter specialists who are representing the final users and who are responsible for preparing the tabulations.

Notwithstanding, at the service center, the error readout is fairly substantial, on the order of 16 to 18 percent for 1980 [8]. Many of these errors are minor keying errors.

Others are editing (abstracting) errors. A number of the nonkeying errors--well over half--are made by taxpayers. Obviously, some so-called "taxpayer errors" are in fact not errors at all, rather they are unusual circumstances which have been erroneously treated as errors. SOI consistency testing may thus introduce mistakes in the data themselves by changing the records in conformance with a model of reality that may not totally fit all situations. Statisticians seem to have an occupational vice; they tend to over-correct or over-smooth (in this context at the microdata level). This predilection, it goes without saying, can be very, very damaging if it is not carefully controlled because it forces on the end user the data model of the data producer and if the data model of the data producer is a very tight model requiring a great deal of consistency in the data relative to the model, then some of the information--maybe even the very thing that is most important--could ultimately be lost. (There is a philosophical issue here about how much error you keep in and how much error you try to take out; someone's error may be someone else's information.) Of course, in addition to possibly "correcting" data that were correct to begin with, error resolution processing may incorrectly "correct" errors made in abstraction and keying.

As a way of controlling the introduction of error into the data due to "overcorrection," a second set of testing is done centrally in the Detroit Data Center for cases where professional judgment has to be brought to bear. Subsequent corrections, at least in terms of the manual adjustments, result typically in one percent of the data being examined or re-examined at this time. At this second round, a great many automatic edits and imputations are also introduced so there are minor impacts on nearly all the items. For example, in the tax system we only balance within tolerances in our administrative processing; for SOI, the balancing is done to the dollar (imputing the out-of-balance portion).

It should be noted here that in recent years, there has been a noticeable decline in the proportion of sample return records that need to be manually corrected in Detroit. For example, the average error rate in Data Center consistency testing for 1973 was a moderate 3.3 percent, but by 1980 it had declined to less than 1 percent.

Three major factors have contributed to these decreases. First, beginning with 1974, data items on the edit sheet were preprinted. The second major factor was the introduction for 1978 of an abbreviated Service Center error resolution. A third factor has been a continuing expansion of the automated consistency tests, permitting the computer to correct an ever larger proportion of the errors in lieu of manual corrections [7].

Other quality assurance activities

Errors, error checks and "corrections" occur at all stages in the processing of SOI. Some of the impacts on quality are measurable,

some are not. In each stage of this system, therefore, there are additional quality review and quality improvement procedures--for example, testing is done of the system itself (systems acceptability testing) and also of the data that are flowing through that system.

In the manual editing (abstraction) of additional information, there is what has been called historically an "edit verification" step at which the manual transcriber's work is reviewed by another person on a sample basis using continuous sampling procedures, which depend on the individual transcriber's error rate (the people who make a small number of errors being reviewed less frequently than the people making a large number of errors).

During key-entry, there is typically a 100 percent key-verification of the data, unlike the manual abstraction which is on a sample basis. This seems, of course, wasteful since much more time is spent reviewing correct data than resolving incorrect data. The SOI Division is examining this processing at this time. But it does unquestionably improve the quality of the data that are produced in SOI. While trying to save money in this budget environment, the Division is trying to find a way to preserve as much of that quality as possible. The already existing computer consistency tests procedures may work to detect most of the mistakes corrected by key-verification. The consistency testing packages though, however carefully they are developed, are not going to identify all of the data capture errors that are made; but nearly all of them may be identified and benefit/cost ratios would suggest that such key-verification could be eliminated or at least done only on a sample basis.

In addition to obvious data capture errors that can be made, there are also errors which are introduced in the estimation side of the system, in the way the data are weighted. One of those was alluded to earlier, in terms of the handling of missing returns. Other problems, of course, have to do with the development of appropriate population totals and their introduction. Errors are made and checks have to be introduced in order to prevent such simple things as data reformatting of the microdata files, and poor table programming, from introducing mistakes into the tables.

Finally, errors of analysis can be made by the end users of the data. Here, of course, is what may be the most fundamental challenge to data producers. In order for the user to properly convert the data supplied to him into meaningful information, we need to provide better measures of quality that are related to his use and also more usable, more readable documentation of what was done and how it is going to affect the user's inference. There has been a widespread failure on the part of data producers to provide the latter.

Generally, the users do not tell the producer what quality they desire, other than an implied "error-free" product which may be impossible or too costly to achieve and in most instances is not needed for the use made of the data. Under these conditions, the producer generally arbitrarily inserts quality measures

into the various processing phases, too much in some, too little in others, with no definite goal and little, if any, knowledge of the quality of the finished product. At IRS, for example, many of our users for some applications could conceivably handle data having less quality than is being given them now, provided there were much better documentation of that quality.

Users of any Federal statistical data must regularly have access to reliable indications of the kinds and levels of error to which those data are subject. Such yardsticks are extremely important since without them users may rely excessively on statistics, presuming an accuracy much greater than actually exists.

2. IMPACT OF DOLLAR BUDGET AND BURDEN BUDGET CUTS ON THE QUALITY OF FEDERAL STATISTICS

The Statistics of Income program, despite its long history and the changes to make it more cost effective and to improve its quality, is subject to many problems affecting statistics in general and statistics from administrative records in particular. Specifically, changes in the amount, kind and quality of statistical information that is collected are directly tied to changes in both the dollar and burden budgets. The implications of these for individual income tax programs will be developed in this section.

In 1981, there were approximately 38 statistical agencies, units, or programs, scattered through more than 90 Federal agencies, producing statistical data and information [9]. The growth in the role of the Federal government since the late 1940's has resulted, until recently, in an increase in the organizations and resources devoted to statistics and to the collection and analysis of governmental data. However, a report by the Congressional Research Service [2] found that between FY 81 and FY 83, the total budget of the Federal government's statistical programs will be reduced by 5.1 percent.

Dollar budget cuts

In March of this year, the House Subcommittee on Census and Population held a day-long hearing to examine the impact of the proposed Reagan budget on the collection and dissemination of statistical information by the Federal government. The testimonies of the witnesses represented the private sector response to Federal statistical data cutbacks. In particular, many of the testimonies focussed on the impact of the budget reductions on the utility and quality of Federal statistics. For example, in written testimony to the Subcommittee, Thomas B. Jabine stated that the quality of Federal statistics is high and that for the most part they are reliable and timely. He notes that now, however, "...the ability of the Federal statistical system to maintain this high standard of performance is seriously threatened." [11] Jabine points out that in the past a significant part of the Federal government's resources for statistical programs was allocated ... to areas that are

of primary importance to the quality of statistics: standards, methods research and development, evaluation and professional staff development." He further emphasized that over the long term, the recent budget reductions, however, are "...certain to result in a serious deterioration of the quality of statistics." Jabine attributes the probable future decline in the relevance and credibility of these data to the fact that "...disproportionately large cuts are occurring in the areas that will determine quality in the future."

In an April 1982 [2] report requested by the Government Operations Committee, the Congressional Research Service analyzed the effects the budget reductions have had on the amount and quality of statistical data available from the Federal government. They pointed out that the budget policies will tend "... to create the most problems for agencies that already use the most efficient procedures because they must reduce the information they collect, process and release." Here the implication of the cuts is different for survey and administrative record statistical systems. Unlike survey systems, often administrative record systems are based on very large samples; sample designs typically are unsophisticated and do not make full use of modern sampling techniques. Often, too, in administrative record systems, again unlike survey systems, automatic editing and imputation are still in their infancy. One of the key strategies for those managing statistical programs using administrative records, therefore, if the quality is not to suffer, is to introduce changes (already made in the survey area) in such a way that the impacts of the dollar budget cuts can be minimized. It may be necessary, for example, to increase the use of sampling for quality control and use greater ingenuity in running the programs which need to be maintained. There may be a need to abandon some programs altogether. For others, a strategy of better quality measurement, but lower quality may be the only alternative. For statistical programs where the quality will have to be maintained or improved, lowering the cost of operations will be the only answer available.

At IRS, the Statistics of Income program has also felt the impact of the budget reductions. A continual erosion of the SOI funding base during the past ten years has resulted in a deterioration in the basic programs to the extent that they are no longer able to fully serve the needs of their major users--the Office of Tax Analysis (OTA), the Joint Committee on Taxation, and the Bureau of Economic Analysis (BEA)--even though every effort continues to be made to offset budget restrictions with increased efficiency in the processing of data.

Congressional budget submissions in support of SOI have steadily declined between FY 1973 and FY 1983. By FY 1981, severe program contractions had become necessary even though significant improvements in productivity had been realized. As a result of the FY 1982 budget reduction, sample sizes had to be reduced considerably, while many data elements were deleted and several programs postponed or

eliminated. For the individual SOI program, the main impact has been to cut the sample to 115,000 returns for 1982 primarily at the expense of State-level detail.

To address this situation, the Commissioner of Internal Revenue and the Deputy Secretary of the Treasury have concluded that non-OTA statistical requests should be funded on a reimbursable basis to the maximum extent possible. For example BEA is going to be asked to fund all of its needs beginning in FY 1984.

Burden budget cuts

The general public is by now quite familiar with the current Administration's dollar budget restrictions. The impact of the Paperwork Reduction Act, however, which imposes "burden budget" cuts on all agencies of government is less well-known. Reductions of this kind are related to the amount of paperwork that each agency requires of the public.

The Federal government first sought to control the growth of reporting burden by the Federal Reports Act of 1942. Over the years, however, the clearance authority became very fragmented, undermining the capacity for coherent burden reduction. In recent years, however, the problems of burden control have been addressed with the establishment of a "burden budget" process (by Executive Order No. 12174 of November 30, 1979), and the enactment of the Paperwork Reduction Act of 1980. The latter completely revised the Federal Reports Act and returned, as of April 1, 1981, all final clearance authority plus statistical policy and standards to the Office of Management and Budget [9]. It establishes a quantitative limit on the amount of paperwork that agencies may require of the private sector or State and local governments in collecting Federal data in order for the government to function. Sizeable cuts are mandated by law; by the end of Fiscal Year 1983 there is supposed to be a reduction in burden of about 25 percent [12].

The extent to which the government is able to manage Federal reporting requirements efficiently and wisely will have a direct impact on the amount and quality of the data produced by the statistical system. James T. Bonnen, in his testimony to the House Subcommittee on Census and Population last March, expressed fears that the burden budget on Federal statistics is "...distorting statistical priorities." He also stated that "...through Fiscal Year 1982 in most statistical agencies, the burden budget has had far more impact on statistical products than the dollar budget. It has also very likely distorted priorities between regulatory and other program agencies and statistical agency collections and uses [12]." A recent paper by the Congressional Research Service cited earlier [2] reported that "...OMB is preparing a new forms clearance guideline that will very likely give greater weight to establishing a Federal need for the data as distinct from broader national needs. Agencies that sought to justify data collection to meet the needs of States, local governments, or private organizations would be required to show that no

alternative was available." There is a brief discussion in [9] of some of the ways in which the management of Federal reporting requirements and the statistical system interrelate. A summary of the main points may shed further light on what the burden budget cuts could mean.

First, the tools used by statistical agencies--sampling, quality control, analysis of existing data, etc.--are near the roots of reporting requirements and can reduce reporting burden if used appropriately. From the point of view of response burden, the use of appropriate statistical techniques is important, therefore, to all Federal data collection. Second, there are numerous major statistical programs which depend on the accuracy and quality of information compiled for administrative and other non-statistical purposes. Administrative records are used both directly (tax return statistics) and indirectly (survey frames for statistical inquiries) in statistical programs. Thus, the level and quality of responses to reporting requirements throughout the government has an important, although often indirect, bearing on the quality of statistical data.

What can be inferred here is that while the burden budget cuts will probably have an impact on the amount and quality of the statistics produced by the Federal government, the impact may be greater on statistics produced from administrative records than those from survey systems. There are a number of reasons that can be given for this. For one thing, the direct burden of the statistical data collection in surveys is very small relative to the total amount of Federal government paperwork (less than 2% in 1981).

On the other hand, cuts in burden can have an important impact on administrative record data. Administrative records systems typically require 100-percent reporting (which is then sampled for statistical purposes in many cases). The burden of the information required is, therefore, a great deal more initially, even though not all that information may be used for a statistical purpose. These cuts may affect the quality of the administrative data collection systems because burden reductions may limit the scope of the data items collected. Careful consideration should be given to the burden reductions so that important data items are not omitted or compromised to the point that user needs cannot be met. Agencies that sponsor administrative data collection must understand the needs of the ultimate user, as well as their own needs, in order to establish proper priorities.

There is a fear, shared by many, that in the course of reducing the administrative paperwork burden, the statistical needs of the society, the need for information, will not be given adequate weight in the decision process. Nor, when in fact it is agreed that certain information should not be collected on a 100-percent basis, will the ingenuity of the statisticians, which is quite impressive in the statistical system as a whole, be brought to bear on devising other methods of collecting that same information on a less than 100-percent basis.

In other words, a way must be found to reduce the burden on the general public; at the same time, the statisticians will have to obtain, possibly through sample methods imbedded within administrative records systems, the information that the society needs. Obviously, if information is not required for an administrative deliberation concerning a particular individual, an economic unit, or a reporting unit, it should not have been collected on a 100-percent basis to begin with. On the other hand, it may need to be collected on a sample basis. The model in the decennial census, of a limited 100-percent enumeration coupled with a more extensive sample data collection, is a model that at least in its general structure needs to be considered as part of the approach to burden budget reductions.

3. OTHER QUALITY ISSUES

There is no question in any of our minds that this is a critical period for the Federal statistical system as a whole. With limited budgets for data preparation and analysis, the coordination of statistical activities with respect to data production in terms of gathering, analyzing, reporting and disseminating and data use will be increasingly important as scarce resources must be distributed among various statistical programs for their maintenance and continued development. Better cooperation is essential both among data producers and between data users and data producers.

To maintain or improve data quality the statistical units or agencies definitely should consider using the challenge of the recent dollar budget cuts as an "opportunity" to rethink ways to share information, and to move towards integrating their systems in areas of mutual concern. Obviously, the present lack of a strong coordinating unit at OMB will make progress in this area difficult. Nevertheless, better cooperation coupled with a focus on more creative approaches to Federal statistical programs can lead to a better integrated system of data collection and may lead to significant savings of resources both for the public in terms of reduced burden and for the government as a whole in terms of reduced cost. A recent proposal for exchange of industry codes between IRS and the Social Security Administration (SSA) by Thomas B. Jabine and Linda Bouchard Taylor [13] exemplifies the kind of ingenuity that may lead us to a resolution of both dollar and burden budget problems without materially sacrificing quality.

Industrial classification of identical or overlapping populations of economic units is carried out by a number of different agencies in the Federal government. A substantial portion of the volume of industry coding takes place at IRS, SSA, the Bureau of Labor Statistics, and the Census Bureau. This has led to problems of comparability for researchers and other users who try to combine or compare data from different agency sources [14]. A look at the IRS and SSA systems in the case study by Jabine and Taylor shows that many of the same

units are being coded by both agencies, and, therefore, there is much duplication of effort. The code-sharing proposal to reduce or eliminate this duplication will require further study, of course, but could result in advantages to both agencies.

As the budget process continues to unfold and limited resources become more of a problem, decisions about quality issues will have to be made with a much stronger focus on the user. As mentioned, user funding is one strategy being explored at IRS so that user needs can be given a sharper focus. For the Bureau of Economic Analysis, this change will put in BEA's direct control decisions about the quality they want.

Challenges to the quality of the Federal statistical data base will continue in a climate beset by reduced funding and severely reduced, if any, central coordination. This should be reason enough, therefore, to work on methods to prevent any further erosion to the quality of Federal statistical information and to find more effective ways of communicating with users concerning their needs for data.

ACKNOWLEDGMENTS

The authors would like to thank Maria Gonzalez and Otto Schwartz for their contributions to this paper. Appreciation is also extended to Bob Wilson for his helpful comments, and to Joyce Coleman, Denise Herbert, Florine Morrison, and, especially, Nancy Robinson for their typing assistance.

REFERENCES

- [1] U.S. Department of Commerce. Office of Federal Statistical Policy and Standards. "Report on Statistical Uses of Administrative Records." Statistical Policy Working Paper 6. 1980.
- [2] Melnick, Daniel, et al. "Recent Changes in the Federal Government's Statistical Programs: An Overview of the President's Budget for FY 1983 and Analysis of the Departments of Energy, Labor and the Bureau of the Census." Congressional Research Service, The Library of Congress. April 8, 1982.
- [3] U.S. Department of the Treasury. Internal Revenue Service. Statistics of Income. 1916-present.
- [4] Blacksin, Jack and Plowden, Ray. "Statistics of Income for Individuals: A Historical Perspective." 1981 American Statistical Association Proceedings, Section on Survey Research Methods.
- [5] U.S. Department of the Treasury. Internal Revenue Service. Estimates of Income Unreported on Individual Income Tax Returns, Publication 1104 (9-79).
- [6] Federal Statistics, Report of the President's Commission, Volumes I and II, 1971.
- [7] Internal Revenue Service, Office of Assistant Commissioner (Planning and Research). "Current Mission and Operations of the Statistics Division." Volumes I and IV. Summer 1980.
- [8] Durkin, Thomas M. and Schwartz, Otto. "The SOI Quality Control Program." 1981 American Statistical Association Proceedings, Section on Survey Research Methods.
- [9] Bonnen, James T., Clemence, Theodore G., Fellegi, Ivan P., Jabine, Thomas B., Kutscher, Ronald E., Roberson, Larry K., and Waite, Charles A. "Improving the Federal Statistical System; Issues and Options." Statistical Reporter, February 1981, Number 81-5, pp. 133-221.
- [10] Casson, John J. "The Deteriorating Quality, Quantity and Utility of Federal Statistics." Testimony before the Subcommittee on Census and Population, Committee on Post Office and Civil Service, House of Representatives, U.S. Congress, March 16, 1982.
- [11] Jabine, Thomas B. Letter to the Honorable Robert Garcia. Subcommittee on Census and Population. Committee on Post Office and Civil Service. House of Representatives. Washington, DC. March 8, 1982.
- [12] Bonnen, James T. "The Need for Better Coordination of Policy and Priorities in Federal Statistics." Testimony before the Subcommittee on Census and Population, Committee on Post Office and Civil Service, House of Representatives, U.S. Congress, March 16, 1982. See also Bonnen, James T. "Federal Statistical Coordination Today: A Disaster or a Disgrace?" 1982 American Statistical Association Proceedings, Social Statistics Section.
- [13] Jabine, Thomas and Taylor, Linda Bouchard, "Sharing Industry Codes: A Case Study." 1981.
- [14] Farrell, Michael G., Jabine, Thomas B., and Konschnik, Carl A. "A Review of Industry Coding Systems," 1982 American Statistical Association Proceedings, Section on Survey Research Methods.