Monitoring Statistics of Income (SOI) Samples

Joseph Koshansky, Internal Revenue Service

For most of its 90-year history, the main function of the Statistics of Income (SOI) Division has been the collection of information for the Department of Treasury and Congress [1]. One of the beneficial practices of a Federal statistical agency, according to the Committee on National Statistics, is its continual development of more useful and timely data, including operational statistics, the latter objective even noted in Internal Revenue Code 6108(a) [2]. SOI has sought ways to improve the quality and timeliness of its tax return information while fulfilling the requests of its primary customers. Over time, it incrementally improved not only the statistical abstraction of information from Federal tax returns, but also the statistical operations associated with producing such information. Moreover, among its various processing tasks, SOI identified the monitoring of its samples of returns from the point of selection to the point of delivery back to the warehouse storage facilities as an essential part of its strategy in achieving its mission.

Because SOI functions within a larger bureaucracy, one of its recurring challenges is coordination among the different staffs laboring at tasks at different phases of the SOI workflow process [3]. For example, in May 2006, the Internal Revenue Service (IRS) awarded a contract to a private company to manage the files function at the IRS submission processing centers [4]. This company will store and maintain all the paper documents taxpayers file at each center for an established period after the completion of IRS "pipeline" processing. It will ship the documents to one of the Federal Records Centers at the end of this period, and fulfill requests from IRS offices that need to examine tax and information returns for either administrative or statistical purposes [5]. SOI is one of the major "downstream" requesters of these stored documents since it produces its mandated annual income, financial, and tax information from weekly samples of Federal tax and information returns, which the IRS usually processes during the previous week [6].

A concern this particular competitive sourcing initiative raises is whether SOI will control within 2 weeks of selection all of the documents in its weekly samples, and not lose some of the returns to other IRS functions requesting *by chance* the same return [7]. On the other hand, the company may introduce new inventory methods or delivery techniques with benefits to SOI, such as interchanges of record information about the pulled returns with one of the SOI databases. Of course, this is not the first time SOI has faced a challenge associated with changes in the way the IRS accepts, controls, and processes tax and information returns. Differences in objectives frequently occur between "pipeline processing" and "postpipeline processing" functions, such as SOI. Ironically, the company will return to an earlier mode of operation SOI replaced through its Total Quality Organization (TQO) initiatives in the early 1990s, shipping "cycles" (or large groups) of returns to the SOI edit sites, instead of program-specific workgroups that SOI units in files supplied to the SOI edit unit editors [8].

This paper is a case study of the infrastructure SOI developed to monitor its samples and deal with unexpected events in a bureaucratic setting. It focuses on what happens after the SOI sampling programs select returns for a project (or study). In addition, it provides an account of the SOI efforts to improve the monitoring of its samples of Federal tax and information returns, part of a "Golden Age" in SOI history. Can regular monitoring of the returns in the various samples decrease the length of time SOI controls returns, or reduce the length of time it finds missing returns in the samples, or reduce the length of time it delivers data to its primary customers? Based on interviews, participant observations, documents, and physical information, the paper shows how SOI operating procedures and information databases, and coordination among different staffs, monitor and verify the control and timely processing of specific sets of returns. In the first section of the paper, we provide a brief historical perspective about SOI consolidation efforts and technological advances. Then, we describe the SOI workflow process in the second section. In the third section, we spell out some of the SOI statistical operations and procedures that systematically monitor the SOI workflow process. The fourth section looks at

the application of management and statistical concepts to the development of the SOI workflow process; and, then, we conclude with several findings and remarks on how SOI is shaping its future.

▶ Consolidation of Work and Technological Advances

SOI performed most of its preliminary statistical abstraction, data transcription, and error correction in National Office, district offices (after World War II for a period of time before the expansion in the number of service centers across the country), and the few service centers in operation, but moved operations to the centers as their number increased. Service centers not only processed but also began storing the paper returns in support of other IRS programs, such as Examination, before final consignment to one of the Federal Records Centers. IRS personnel at the different SOI sites, who were available to edit SOI samples once regular pipeline processing work subsided or ended, used paper edit and error register sheets to abstract information from the returns, while National Office analysts produced aggregate statistics and tables from the perfected data for customers [9].

In the 1980s, under the direction of Fritz Scheuren, SOI adopted the Total Quality Organization (TQO) methodology to improve its operations at the service centers and in National Office, primarily in response to a request from analysts in the Office of Tax Analysis (OTA) and Joint Committee on Taxation (JCT) for *earlier* deliveries of SOI data. SOI analysts identified vital activities and formed cross-functional teams to work on these issues. The staffs in the different branches in SOI National Office looked for ways to develop work processes and data systems that could improve the quality and timeliness of the tax return information they produced for each of the SOI programs within the boundaries of regular IRS pipeline processing. The research included traveling to the service centers to meet with employees for the purpose of identifying, prioritizing, and recommending improvements in SOI control and processing of returns in its various samples [10]. According to Scheuren, "[t]he focus on process quality that Deming and Juran urge, while not really new, is having a revolutionary impact on us, especially in its emphasis on continuous improvement or "Kaizen," as the Japanese call it…. Examples [include] more flexible and dynamic approaches to data capture, cleaning, and completion" [11].

From this analysis, Scheuren and others on his staff hypothesized that consolidating SOI editing operations at particular IRS service centers would free up resources (staffing, travel, and training), improve editing (abstraction) productivity and quality, and enhance its presence as a data producer within the community of Federal statistical agencies. In May 1990, SOI notified the now ten IRS service centers that it planned to consolidate edit processing for the SOI Corporation and Individual Tax Return programs in six service centers [12]. Four centers would only pull, control, and ship returns to one or more of the six processing centers (down to five in 1992) [13]. In general, the number of returns service centers processed for all of the SOI studies was much smaller than the volume of returns the centers processed for tax liability, administrative, and informational purposes. Competing with other functions for skilled tax examiners to work the SOI programs at the centers, as well as arguing about what IRS or SOI programs merited attention first, were frequent occurrences before the consolidation initiative.

Concentrating the editing function at six service centers led to the formation of additional units of SOI editors (former tax examiners and data transcribers) at some of these sites and the growth in the volume of available work at all the sites [14]. Most of these edit units were now dedicated to processing only the returns in SOI samples year round. SOI ensured the volume in each of the six processing centers was sufficient to support an SOI edit unit working full-time on SOI work. Besides the formation of SOI edit units, SOI created "SOI control units," at least in name, in each of the ten centers' files warehouses to support its edit units. After regular pipeline processing, each of the centers stored for about 2 years its portion of the total population of returns that filers mailed each year. An SOI control unit consisted of a small group of service center employees, usually working in a miscellaneous unit in the files, whose major tasks were the control, processing, and shipping of returns in SOI samples to the SOI edit units and refiling returns after edit units completed pro-

cessing these returns. SOI discovered a truly dedicated group of employees, who shared their files expertise and experience in searching for and finding missing returns, as well as assisting National Office analysts in finding additional information about certain returns [15].

While one National Office cross-functional team was working on the consolidation initiative, other teams were developing new online computer applications and installing new hardware at the centers, solely dedicated to SOI processing. Beginning in 1991, SOI procured and installed hardware upgrades and telecommunication equipment for support of online editing, at the Cincinnati and Ogden service centers, and in National Office. Telecommunication lines connected online terminals for the editors in each of the processing centers to the SOI minicomputers in Cincinnati and Ogden, designated SOI minicomputer hub sites. The integration of editing, data transcription, and error correction into a single operation with these online terminals began with several smaller SOI studies (Partnerships, Exempt Organizations, Controlled Foreign Corporations, Foreign Tax Credit, and Individual Sales of Capital Assets) and expanded to the major Corporation and Individual Returns programs. Online editing brought significant improvements in productivity, timeliness, and quality because editors spent much less time waiting for nightly batch-mode feedback on errors and corrections and much more time processing completely sets of the same type of return [16]. Groups of tax examiners became experienced subject-matter experts on how filers completed forms, as well as knowledgeable about the content of the forms in question. Having honed their skills from frequent and consistent editing of a large number of the same type of return, they accelerated processing and improved the quality of the final product—perfected and more meaningful return information [17].

The availability of returns to edit on a continuous flow basis was an important concern now that service centers increased the size of their SOI edit staffs, and in some cases improved the grade structure, to deal with the increase in the volume of work. Would the edit units have enough work? Would the editors' work habits outpace the delivery of new returns to process? Would waiting for work adversely affect the earlier training and skill levels of the editors? Managers in the SOI edit units identified one of the requirements for successful execution of the new plan as timely delivery of a sufficient amount of returns. Timely delivery of work supported the efforts of centers to commit employees to SOI projects the entire year, *so long as SOI work was available*. Consequently, another National Office team developed an online database application, called the SOI Automated Control System (SOIACS), to monitor, first the shipment of 1040 returns, then all returns [18]. A next-generation version of the application, now named STARTS, would facilitate the "systematic control" of 1040 returns some service centers would ship to other centers for edit processing, as well as the movement of returns between an edit unit and control unit within the same center [19]. Subsequently, when operational, the application had a computer terminal and printer located in the files of each of the ten service centers and the edit units [20]. It connected the control units with the edit units and both with National Office.

Soon after implementation of the application, an edit unit manager's need to know what returns to edit first (i.e., the editing priority) surpassed the need for timely delivery of returns because SOI began committing to deliver data to its customers by specific dates during the year. The centers needed meaningful information to answer this and other questions. For example, a question an SOI edit unit manager might raise is, "Which returns in the cycle (weekly pull) should we process first?" But a new SOI files clerk might ask, "If another IRS function has the return, can I pick another one on the same shelf (for SOI)?" SOI editors might ask, "What returns do I edit?" or "Where do I move this money amount?" An SOI National Office statistician might ask, "Can we ask the centers to locate the missing returns?" An SOI economist might ask, "Can the centers edit more of the Type XYZ returns (for example, Sample Code 20 or Cross-Sectional returns) before the deadline?" Finally, an SOI scanner might ask, "How do I replace the illegible page?" These questions demanded better monitoring not only of the physical location of the returns while en route to the edit units, but also better visualization of the metainformation of the returns—i.e., information that describes the information about a sampled return [21]. Now that SOI created an IT backbone to support its workflow process, managers asked for more details about what actually was in a cycle of returns [22].

▶ SOI Workflow Process

Compared to IRS administrative processing, which captures some information from all of the filed tax returns, SOI studies collect much more information from samples of returns through its transcription and editing. SOI editors add value to the administrative record information the IRS collects. This additional value makes it imperative to control and monitor the samples and continuously improve the entire SOI workflow process to guarantee consistency over time. Similarly, information about the processing tasks adds value to the corresponding returns that flow through the workflow process. The results of the efforts of the TQO teams in collecting information at each phase of the process about the processing tasks; the performers of these tasks; the relative order of the tasks; the possible synchronization of some of the tasks; the flow of information in support of the tasks; and the tracking of the tasks, was not only a better understanding of the process, but also a cache of aggregated information.

The SOI workflow process is the general term for the movement of samples of "documents" or "containers of information" (e.g., paper returns, electronic records, and digitized images), through the SOI sampling, controlling, and editing processes [23]. Each of these three major subprocesses, or phases, relate to specific tasks that personnel at the service centers and in National Office execute to produce statistics for publication and delivery to customers. Both operating procedures and computer systems support the efforts of the people involved at each of the phases of the process. This convergence of procedures, databases, and people forms an underlying base, or infrastructure, for the functioning of the workflow process.

The process begins when a project analyst adds a new tax or information form to an existing study or initiates a new study with an SOI customer. After the SOI sampling programs at the IRS computing center, or the Ogden Submission Processing Center, selects returns for a particular study, the programs then create sets of output files for loading into both IRS and SOI databases [24]. Phases of the process include selecting documents, pulling documents, monitoring the success rate of pulling documents, finding missing returns, storing documents, scanning documents, photocopying documents, ordering documents, shipping documents, editing documents, managing documents in the edit unit, and releasing documents back to files. The process involves constant change and update. For example, under the new competitive sourcing initiative, the SOI edit units at the centers will assume tasks the SOI control units once performed after the contractor begins managing the Files function at the centers. The infrastructure alleviates some of the problems associated with such a change.

▶ SOI Monitoring Operations

The Statistics of Income Automated Return Tracking System (STARTS) is the framework for management of returns and digitized records as they move through the various phases of the SOI workflow process at the centers. This process control system is a structured set of related components (people, procedures, processes, subsystems, databases, reports, etc.) SOI established to accomplish the major task of monitoring its samples from the point of selection to the point of delivery back to files. STARTS (the system) consists of online database applications, as well as standardized business processes, work instructions, forms, and reports, all of which give the different staffs at the centers and in National Office increased visibility into the operations at the centers.

The SOI sampling program, sample selection sheets, document chargeout forms, pulled returns, shelved returns, and shipped workgroups of returns, comprise part of a "signal" system for securing and delivering the correct returns in an SOI sample to the right service center for processing at the right time. The other part is the database, developed for predictable and manageable record keeping.

Database Management System

Borrowing from manufacturing operations, which schedule and track the flow of materials through a process, STARTS (the database application) gives online access to real-time data about one return, or a group of returns (cycles, workgroups, scanned sets, photocopied sets, etc.). Combining aspects of transaction processing, management information, decision support, and

expert systems, the database is a collection of information about SOI samples, which users manage and utilize when making decisions about planning, organizing, and controlling the processing of the samples [25]. Top-level managers are concerned with planning: Will the center meet the corporation program 75-percent cutoff on the scheduled date? Middle-level managers are concerned with organizing: Can the editors in Unit 5 handle the consolidated 1120 returns? Front-line managers are concerned with controlling: Are the editors; documents, scanned images, or electronic records; and inventory and edit applications available to begin editing the corporation returns?

Convergence of Aggregated Information

Because STARTS (the database application) stores sample information and provides a traceable record of user transactions or interchanges with that information, one example of its functioning is worth noting here. A section of the Internal Revenue Manual (IRM) notes the date the centers must supply transcribed and edited 1040 return information to National Office for "Advance Data" delivery to OTA and JCT. One year earlier, mathematical statisticians produced the sampling specifications for the computer specialists who wrote the programs that selected returns for the sample. Among the possible inputs, the application reads and stores return information that the sampling program at the IRS computing center loaded into the SOI sample control files, or the "One-Week Followup" date a clerk entered in the STARTS cycle control screen. The application applies a set of logic statements (or SOI business rules) to the loaded records, such as, if the Level Code is equal to "1," or the Continuous Work History Study (CWHS) Code is equal to "1," assign the return to the "Cross-Sectional" category, or if the sample code of that return is a specific value within a certain range, assign it, as well, to an additional category, called "Complex" edit. Possible outputs include the application generating and displaying inventory totals, such as the number of "Complex Cross-Sectional" returns, which are available for the SOI edit unit manager to order, or permitting the placement of a user-defined set of these "Complex Cross-Sectional" returns into a STARTS editor workgroup.

▶ Application of Management and Statistical Concepts

A "Golden Age of SOI Development" occurred at the end of the 1980s and the beginning of the 1990s in SOI National Office and the centers, which resulted in an infrastructure that is still in place today. Inhouse "quality" teams of economists, management and program analysts, statisticians, center managers, editors, clerks, and information technology specialists collaborated in the design, development, application, and maintenance of this infrastructure. Based on the research of American experts such as Frederick Winslow Taylor, Frank Bunker Gilbreth, Walter Shewhart, and of the War Department's Training Within Industry, SOI learned that continuous incremental improvements benefit an organization [26].

Value

SOI increased the value of the tax returns in its samples not only for its customers, but also for its suppliers at the service centers (see Table 1).

Table 1—Added Value At Each Phase of Workflow

Pull and control documents	Document information Location information Cycle information Pull information
Store documents	Warehouse information Center information Time information Processing information
Order and ship documents	Return information Project information Edit priority information Edit site information Workgroup information Center information Complexity information Deadline information
Process documents	Edit information Scan information Photocopy information Critical case information Split-screen information
Release documents	Quality review information Refiling information

SOI assigned information, based on descriptive statistics from different operational sources, to each return record to expedite processing. Identifying and storing information about a return, its edit status, and its extra-processing requirements in a database made the fulfillment of requests for any of this information much easier. For example, the set of all possible outcomes of an operation at a particular phase of the process determined whether a return was released immediately after editing, instead of scanned. Consequently, a supply chain concept replaced the original "shipping" concept. The SOI infrastructure moved not only documents, electronic records, or digitized images, but also information from unit to unit, center to center, headquarters to field office.

Complexity

The purpose of the process control system shifted from one where the principal activity is moving documents from one center to another to one where the activity is helping the centers meet the program completion deadlines, which National Office analysts set to provide timely tax return data to its customers. SOI managed complexity, sometimes even reducing it, when it assigned returns in the various project samples to a series of categories. Combinations of these categories made it possible for the managers to break down the amorphous cycles of returns into pieces that are easier to control and work with. Since it is necessary to edit some returns before others, the STARTS application provided the capability to order specific sets of returns, placing them in specific sets of editor workgroups. These combinations supplemented the strata the math statisticians created for sampling.

Standardization

The STARTS application allows SOI to standardize certain processing tasks across the projects and the service centers. It acts as a decoder that helps personnel in National Office, the SOI edit units, and the SOI control units to understand each other's variants of sample processing. The corresponding system makes these different actors work together through the interchange of information. They have to follow certain rules to avoid miscommunication and guarantee that both the SOI edit units and SOI control units know in advance from the information in the database application what each should provide as updates or requests and what each should expect back as responses. When an edit unit orders 20 editor workgroups in which each workgroup contains ten "Priority 1" corporation returns, it expects the SOI control unit to assemble and send 200 such returns for distribution to five editors. Because the SOI control unit marks a return as "missing" in STARTS if it does not control that return, only what is in its control is available for the SOI edit unit to order in STARTS.

Kaizen

The consolidation efforts changed SOI into an organization that continues to apply time-compressed, action-oriented improvement methods to its various projects. Many of the components and functions of the STARTS application were the result of the energy generated through users' participation, creativity, and the pressure to produce rapidly tangible results.

► Conclusion

The formation of cross-functional teams at the centers, and between the centers and National Office, and the development of a monitoring system and corresponding just-in-time electronic database application (i.e., STARTS) brought a very strong focus on the entire SOI workflow process. No function could make a change that affected another function unless they had buy-in from that function. Managers, editors, clerks, statisticians, economists, analysts, and computer specialists looked at samples from beginning to end, not just a particular phase. The teams monitored the status of returns as they "flowed" through the workflow process.

When the private company begins managing the IRS files warehouses at the centers in late 2006 and sends the first batch of pulled returns to the SOI edit units, days before the arrival, SOI National Office and its SOI edit units across the country will know what returns the SOI sampling programs selected for the various studies. Unfortunately, the company will not exchange electronic records with STARTS per the contract. In addition, SOI will no longer have a presence

in the files warehouses per the IRS performance work statement. SOI personnel both in National Office and at the edit units at the centers will not know the contents of the shipments until the SOI edit units can open the boxes or scan the carts. If the company transmitted an electronic version of the shipment manifest for loading into the STARTS database application, then the SOI edit units might consider shelving the returns in workgroups for easy distribution to the editors, instead of storing in a traditional files manner (e.g., cycle or type of return).

In the future, if an SOI edit unit runs low on work, the STARTS database application could recognize this situation in the inventory and order more. Because this application stores record information for each return in the sample, whether processed as paper, an electronic record, or a digitized image, SOI can easily repurpose the record content, making it accessible from a variety of devices.

The database application increased the availability and use of data, consequently helping to improve each center's decisionmaking and visualize, synchronize, and automate phases of the workflow process. The power in STARTS reports and screens is that they display accurate, consistent, and timely data. SOI built a reporting system so that managers know in real time how they are meeting the needs of SOI customers. The application replaced transactions done by phone, fax, or mail. It replaced collecting and storing data manually in their own way.

In the late 1980s, SOI developed online data entry and verification applications, which linked IRS processing sites across the country through a network of computer terminals and databases. It applied this information network concept to the control and monitoring of its samples. This connectivity and the value-added information embedded in each sample record allowed SOI personnel to monitor the status of each tax and information return as it moved through the different phases of the SOI workflow process from the files warehouses to its edit units and back. Incorporating a wide range of information about the sampling criteria, the study objectives and requirements, and the logistical demands associated with processing enhanced the meaning of the samples to the centers (suppliers) and National Office analysts (producers) and assured an acceleration of the collection of data and the delivery of the final products to SOI customers. Monitoring daily the number of missing and available returns can increase the likelihood the quality of the data is high [27].

▶ Acknowledgments

The views expressed in this paper represent the opinions and conclusions of the author and do not necessarily represent those of the Internal Revenue Service. The author thanks John Czajka for his comments of an earlier draft of this paper. Any errors that remain are the responsibility of the author.

▶ Endnotes

[1] In addition to the Office of Tax Analysis and the Joint Committee on Taxation, another important customer is the Bureau of Economic Analysis.

[2] National Research Council (2005), *Principles and Practices for a Federal Statistical Agency, Third Edition*, Committee on National Statistics, Margaret E. Martin, Miron L. Straf, and Constance F. Citro, editors, Division of Behavioral and Social Sciences and Education, The National Academies Press, Washington, DC, p. 25. In addition, see 26 USC Sec. 6108, Statistical publications and studies, which describes the SOI mandate.

[3] The SOI workflow process is the interchange of documents, record information, and tasks through the SOI sampling, controlling, and editing processes.

[4] As a stakeholder and customer, SOI hopes to meet with company representatives and the IRS Files Government Project Management Office to discuss pertinent issues about its samples. After announcing the awarding of the contract, the IRS announced two positions, one a senior manager position, the other a supervisory quality assurance specialist. While a company assumed responsibility for the work performed in files, it

is necessary to manage the relationship between this company and other IRS offices and check the quality of the company's work, etc.

[5] The company will operate at the IRS facilities in Methuen, MA, Fresno, CA, Norcross, GA, Austin, TX, Ogden, UT, Kansas City, MO, Florence, KY, and Philadelphia, PA. The records centers are part of the National Archives and Records Administration. They store the records of a Federal agency.

[6] In addition, SOI is a major requester of electronic records, which include electronically-filed records.

[7] Competitors for documents include four different business operating divisions: Large and Mid-Size Business (LMSB), Small Business/ Self-Employed (SB/SE), Wage and Investment (W&I), and Tax-Exempt and Government Entities (TEGE).

[8] The acronym "TQO" refers to Total Quality Organization, a commitment on the part of an organization to advocate quality and continuous improvement in all its tasks.

[9] The general term, "regular pipeline processing," refers to the actions of IRS workers who handle tax and information returns from the time the documents first arrive at an IRS service center through the posting of information at the IRS Computing Center and finally the shelving of the documents in the files area.

[10] SOI wove supplier and customer data into the process improvements. It captured any available information relevant to the SOI projects at the centers.

[11] Scheuren, F. (1991), Comment on "The Federal Statistical System's Response to Emerging Data Needs" by Jack E. Triplett, *Journal of Economic and Social Measurement*, IOS Press, Volume 17, Numbers 3, 4, p. 190.

[12] The 1990 plan for distributing work to the remaining six processing centers had Andover and Brookhaven centers shipping their individual and corporation returns to the center in Ogden. Memphis shipped its individual returns to the Austin center and corporation returns to the center in Cincinnati. Philadelphia shipped both individual and corporation returns to Cincinnati. The Atlanta, Fresno, and Kansas City centers continued to process their samples of individual and corporation returns. Doug Shearer and Dan Trevors coordinated the plans and issued regular status reports to keep management informed of the activities involved in this consolidation. For the Individual program, the consolidation was effective beginning with the Cycle 9053 End-of-Year Tickler (EOYTICK) processing for the Tax Year (TY) l989 Study and continued with the TY 1990 Study, which began with the selection of returns in Martinsburg Computing Center (MCC) Cycle 9104 (January 1991). Consolidation of the Corporation program began earlier with the TY 1989 study commencing only in Atlanta, Austin, Cincinnati, Fresno, Kansas City, and Ogden in August 1990. The nonprocessing centers began shipping their corporation returns to the edit sites later in the year per SOI notification. Beginning in 1992, the edit processing of the returns in the Individual and Corporation programs resided in only five centers, when SOI discontinued editing at the Fresno center.

[13] The centers were located in Andover, MA, Brookhaven, NY, Memphis, TN, and Philadelphia, PA. A team of managers from National Office traveled to these centers to discuss issues and concerns of the managers, editors, and clerks.

[14] SOI editors abstracted information from returns, including moving some information to the correct fields on the returns. Tax examiners in non-SOI units at the centers checked and prepared for data transcribing those fields on the returns the IRS deemed important in determining tax liability.

[15] Clerks in the SOI control units did not edit returns. Instead, they pulled returns, looked for missing returns, photocopied returns, scanned returns, packaged returns, and shipped returns to list just some of their duties. One manager commented: "I am a Unit Supervisor in a large unit. I have IMF SOI, AIMS, Cycle, Quality Review … as well as pulling and refiling. SOI is just a part of this unit. We have maintained a record of high accuracy and very few missing documents for a few years. This [is] … due to the integrity, dependability, and dedication of the staff assigned to SOI. They have accomplished a lot with very few people. So, what STARTS means to me is reflected in what the staff commented on … If they are happy and satisfied and feel that STARTS helps them perform their duties more efficiently and accurately due to the increased speed and easier access, then I am happy. If they feel that STARTS helps them maintain a low missing record, and this record is reflected on the SOI reports for Andover, then I am happy with STARTS. I do not use STARTS myself, but I do review the reports that these employees generate."

[16] Editors usually waited the next day to receive feedback because centers scheduled SOI batch programs around regular pipeline batch jobs.

[17] It is difficult for an editor to maintain his or her skill level if he or she moves frequently from one project to another, though the frequent changes may guarantee work for that employee.

[18] The developers considered SOIACS the first step in building a system to manage its samples in an online environment. SOI planned to build subsystems to manage quality, resources, and sample selection as part of the modernization effort because the service center statisticians were retiring or service center management considered them irrelevant. Dan Trevors of the Quality Support Team and Doug Shearer of the Coordination Team shared responsibility for developing the SOI controlling and shipping process. Linda Taylor of the Distributed Processing System Team provided hardware support. The SOI operating branches, as well as the service center files and edit operations, defined, collected, and presented the user requirements. A manager's comment: "The STARTS system is a valuable tool used on a daily basis. It helps track the work … as well as when it is edited within the edit teams. When a return is marked missing and we find it attached to another return, we are able to go to the remarks [screen] at that time to document the condition. The STARTS system is also used to look up prior-year information. If an EIN is the only information you have to track component parts of a separated 1504C return, the STARTS system can provide much information on this. This helps us to locate additional return parts in order to edit a more complete document. STARTS provides many options in ordering the work. It is broken down by return type, three asset class categories, and the sample code only selection of returns. This gives management the necessary range to order specific types of work at all times but is especially helpful when nearing various project completion dates. As transition continues here in Ogden, we are very interested in the future STARTS process and the new and evolving ways in which we will utilize the system. We look forward to the changes and future training that is available to all leads as well as the clerks and managers."

[19] National Office analysts held a planning session with service center personnel the week of June 18, 1990, at the Austin Service Center to collect ideas, customer needs, and specific requirements for the SOI Automated Control System (SOIACS). Back in National Office, the team reviewed the requirements, analyzed the consequences of implementing a control system, and wrote descriptive and detailed requirements and specifications, which bridged the requirements and the design of the application. Cincinnati Service Center assumed primary responsibility for the Oracle program development of this new application, with Don Flynn as the lead programmer. Tentative plans involved piloting the application in one processing center and one

nonprocessing center in the spring of 1991 for the Individual returns project. The SOI programming staffs at the Cincinnati and Ogden Service Centers developed the next generation of the application, which National Office renamed the Statistics of Income Automated Return Tracking System (STARTS). The Cincinnati staff developed and maintained the Individual Master File (IMF) version of STARTS, while the Ogden staff programmed and supported the Business Master File (BMF) version. In 2000, both programming staffs converted the text-based applications to a graphical user interface (GUI) application.

[20] Connections between the center terminals and the host minicomputer in Cincinnati occurred through PACNET.

[21] In the case of tax returns in SOI samples, this is metainformation about relational database properties; data warehousing; business intelligence; general IT; IT metadata management; file systems; and image, program, project, and study schedules.

[22] SOI assigned information to each return: project, sample, files location, edit site, editor, delivery dates, level of edit complexity, document source (paper, electronic, or image). One result was a sample redesign, which embedded a panel within the annual cross-sectional samples. The STARTS application still distinguishes these two sets of returns. See Czajka, J. and Walker, B. (1990), Combining Panel and Cross-Sectional Selection in an Annual Sample of Tax Returns, *1989 Proceedings of the American Statistical Association, Section on Survey Research Methods*.

[23] The use of digital images, instead of paper, as source documents for editing is a new phase in the SOI workflow process. Other SOI processes include data cleaning and completion, weighting and estimation, and publishing tables and user analyses.

[24] Systems acceptability testing (SAT) occurs before the computing centers execute the SOI sampling programs. Sample design and sample selection are topics for further discussion in other papers.

[25] Stair, R.M. (1992), Principles of Information Systems: A Managerial Approach, Boyd and Fraser Publishing Company, Boston.

[26] Maurer, R. (2004), One Small Step Can Change Your Life: The Kaizen Way, Workman Publishing Company, New York.

[27] Improving data quality through editing, imputation, and record linkage is impossible if the administrative records that contain the data are unavailable or incomprehensible.