Statistics of Income – Distributed Processing System (SOI–DPS) – Privacy Impact Assessment

PIA Approval Date –   November 2, 2011

System Overview:
The Statistics of Income Distributed Processing System (SOI–DPS) is a geographically distributed processing environment composed of midrange computers, workstations, and thin clients and is essential to the success of the SOI program. SOI–DPS is designed to facilitate the collection, production, and publication of statistical data for use in the formulation and measurement of legislation relating to taxation as required under IRC 6108. SOI–DPS does not receive data from Federal, State, or Local Agencies. SOI utilizes external data sources such as Mergent Online to perfect data for statistical purposes. These data are based on individual, corporate, partnership, exempt organization, and estate tax returns, plus those related to foreign activities.

Systems of Records Notice (SORN):
- IRS 34.037--Audit Trail and Security Records
- IRS 70.001--Individual Income Tax Returns, Statistics of Income

Data in the System

1. Describe the information (data elements and fields) available in the system in the following categories:

A. Taxpayer – Data are extracted from almost every line of the returns designated for SOI studies:
- Name
- Social Security Number (SSN)/ Taxpayer Identification Number (TIN)
- Address
- Wages
- Withholdings

The data are perfected and become part of microdata files on which the SOI estimates are based.

B. Employee – It is possible that as taxpayers, some employees' returns may meet the criteria for inclusion in SOI samples. Data elements include:
- Name
- Employer Identification Number (EIN)
- Address
- Industry Codes

C. Audit Trail – The System Administrators (SAs) /Database Administrators (DBAs) make periodic activity log file checks to ensure normal system functioning and to watch for warning messages that might indicate a problem. The activity log contains:
- Logon and logoff of all users by USERID
- Password change including USERID
- Date and time of event
- Success or failure of the event
- All actions by System and Security Administrators
- All actions by DBAs

2. Describe/identify which data elements are obtained from files, databases, individuals, or any other sources.

A. IRS – Taxpayer data comes from the following systems:
- Service Center Recognition/Image Processing System (SCRIPS) – SOI–DPS receives K1 image (name, SSN, address, other tax return data) and data files from SCRIPS at the Cincinnati and Ogden hosting and editing sites.

- Modernized Tax Return Database (MTRDB) [Subsystem of Modernized e–File (MeF)] – SOI–DPS receives data files (name, SSN, address, other tax return data) from MTRDB at the Ogden hosting and editing site.

- Individual Master File (IMF)/Business Master File (BMF) – SOI–DPS receives data files (name, SSN, address, other tax return data) from Master File at the Cincinnati and Ogden hosting and editing sites.

- Large Business and International (LB&I) Data Capture System (DCS) – SOI–DPS sends images of tax returns (name, SSN/TIN, address, other tax return data) that were scanned to the DCS application.

B. Taxpayer – Data are extracted from the Individual and Business Master File systems for sampled returns representing virtually every type of tax and information return filed with the Service. Paper returns or images of scanned returns may be used in this process. The data elements for any given SOI study are defined in the processing documents maintained by the SOI program branches.

C. Employee – It is possible that as taxpayers, some employees' returns may meet the criteria for inclusion in SOI samples. Data elements include:
- Name
- EIN
- Address
- Industry Codes

3. Is each data item required for the business purpose of the system? Explain.
Yes, each data item is required for the business purpose of the system. SOI–DPS is designed to facilitate the collection, production, and publication of statistical data for use by customers in the formulation and measurement of legislation relating to taxation as required under IRC 6108. This statistical data is based on tax and information returns, most of which are filed annually.

4. How will each data item be verified for accuracy, timeliness, and completeness?
The SOI–DPS system checks for accuracy, timeliness, and completeness of the information inputs.

- **Accuracy** – The "raw data" that is used by the SOI–DPS applications is the taxpayer data that is provided by the taxpayer on their tax return. SOI–DPS runs customized programs to select the data in accordance with the specifications provided by the customer for a given study. The Editor forms (written in Oracle) are pre–populated with this data. Therefore, the data is assumed to be accurate as it is provided by the taxpayer. Other fields are manually input by the Editor using the original tax return data. Format and syntax checks are in place for these fields. If the required fields are not completed correctly, the application prevents the user from moving to the next screen. SOI employees use various mathematical and validation methods

to check selected data for accuracy. As part of this process, actual returns or images of scanned returns may be used to verify the accuracy of the data that is input.

- **Timeliness** – Data are based on current year filings. Most SOI studies are annual. SOI–DPS data retained in the system are static once the study files are completed.

- **Completeness** – Each screen has a defined set of required fields. If the user does not complete these required fields with accurate information, the user is prevented from continuing to the next screen.

The data for selected returns are input into the Oracle database resident on the SOI systems in the Cincinnati and Ogden Submission Processing Centers. Additional data are extracted from the physical tax return or an image of a scanned return to prepare a complete record for SOI. After subjecting the data to various consistency and validity checks, records are input for further statistical processing on SOI's systems in National Office in order to produce data for transmittal to SOI's customers.

5. Is there another source for the data? Explain how that source is or is not used.
No. SOI–DPS data are only collected from selected tax and information return forms in accordance with legislative mandates. Data are extracted from the Individual and Business Master File systems for sampled returns. SOI–DPS receives data from different systems – e.g., the images and associated data from written IRS Forms K–1 that are scanned in the SCRIPS system, and the XML data from Modernized Tax Return Database (MTRDB). Except in very specialized circumstances, there is no interface between the IRS and the taxpayer to obtain any additional information. Liaisons from the eFile project office are used to interface the taxpayer to the IRS. Taxpayers may also be contacted via the mail by SOI to obtain missing tax information for statistical purposes only, but are under no obligation to provide the data to SOI.

6. Generally, how will data be retrieved by the user?
Users are given access to certain data sets, files, and images based on permission and user roles. Users log into the system and can view the data their managers give them access to. Queries are performed by EIN, SSN, organization name, or other data element depending on the imaging subsystem being used.

7. Is the data retrievable by a personal identifier such as name, SSN, or other unique identifier?
Yes. Individuals with the proper permissions can access the data and images of returns using Social Security Numbers, Employee Identification Numbers, Taxpayer Identification Numbers, Industry Codes and any other information contained in the database for that return. Users can only pull data for reports they are working on.

Access to the Data

8. Who will have access to the data in the system (Users, Managers, System Administrators, Developers, Others)?

Role: SOI–DPS Editors at Service Processing Centers (SPCs)
Permission: Extract data from various tax returns

Role: SOI–DPS Analysts at National Office
Permission: Economists and Math Statisticians working on particular studies

Role: SOI–DPS Managers
Permission: Manager of various projects that control access to the data for a particular SOI Study

Role: SOI–DPS Developers and Testers
Permission: Develop and Test SOI–DPS Application

Role: SOI–DPS System and Database Administrators
Permission: Maintain SOI–DPS hardware and databases

Role: Treasury's Office of Tax Analysis (OTA)
Permission: Use microdata for various SOI Studies for tax analysis

Role: Congressional Joint Committee on Taxation (JCT)
Permission: Use microdata for various SOI Studies for tax analysis

Role: Bureau of Economic Analysis (BEA) at Commerce Department
Permission: Use microdata for various SOI Studies for analysis. No login to SOI–DPS

Role: Private–sector customers (Urban Institute, Philanthropic Research Institute, Foundation Centers)
Permission: Access to images of certain publicly available tax returns for private use. No login to SOI–DPS

Note: All contractors who access SOI–DPS have appropriate IRS clearances and/or background checks.

Permissions/access is dependent on the role and the imaging subsystem the user is working on. Please see below for more details.

- The Imaging Subsystems allow the Editors and Viewers to edit and view tax return images, respectively. Access to the SOI–DPS Imaging Subsystems is restricted as follows:

  o SOI Exempt Organization Image Net (SEIN). Access to SEIN is restricted to SEIN Viewers, SEIN Editors and SEIN System Administrators. The Viewers can only view the tax return images. The Editors can view and edit the tax return images. The System Administrators (located in Ogden) are responsible for the imaging of the tax returns and have administrator–level access to SEIN.

  o Document Image Net (DIN). Access to DIN is restricted to DIN Viewers and the DIN System Administrator. The Viewers can only view the tax return images. The System Administrator (located in Ogden) is responsible for the imaging of the tax returns and has administrator–level access to DIN. In addition, Treasury has view–only access to DIN.

  o Large Business and International (LB&I) Image Net (LIN). Access to LIN is restricted to LIN Viewers, Powerusers, and LIN System Administrators. The Viewers can only view the tax return images they are authorized to view and/or download them (XML data only) to an SBU directory. The Viewers receive an email message with a link to the tax return image. The Powerusers have access to create folders and images. The System Administrators have access to alter customer/user profiles and have administrator–level access to LIN.

- o Paper Image Net (PIN). Access to PIN is restricted to PIN System Administrators, a Scanner, and several Document Preparers. The PIN subsystem is not accessible by the customer. The customers send in paper documents to the PIN System Administrators. The documents are scanned in the order they are received, put on encrypted CDs, and sent to the customer. The paper documents are either destroyed are sent back to the customer. The Document Preparers prepare the documents for scanning (i.e. pulling staples). The Scanner has access to PIN to scan the images onto the server. The System Administrators are responsible for placing the scanned images of the documents on cd and have administrator–level access to PIN.

- o K1 Image Net (KIN). Access to KIN is restricted to Viewers, other LB&I users, and the KIN System Administrator. The Viewers can only view the K1 forms images. Other LB&I users can view, query and print the K1 form images. The System Administrator is responsible for the imaging of the K1 forms and has administrator–level access to KIN.

Note: Viewers are located all over the U.S.; including Guam, Puerto Rico, and the Virgin Islands.

- The Imaging systems listed above are supported by Oracle forms. The Oracle forms are completed by Editors to populate the SOI–DPS data file, which resides on the SOI–DPS database server. The Oracle forms are the back–end of the SOI–DPS data processing process. The data file is read by the Statistical Analysis Software (SAS) program, formatted, and queried, to produce statistical results based on requirements of customer studies.

9. How is access to the data by a user determined and by whom?
Access to the data is determined by the manager based on a user's position and need–to–know. The manager will request a user be added. They must fill out an Online 5081 (OL5081), Information System User Registration/Change Request, to request access to the application. JCT users use the paper Form 5081. A user's access to the data terminates when it is no longer required. Criteria, procedures, controls, and responsibilities regarding access are documented in the Information Systems Security Rules within OL5081.

10. Do other IRS systems provide, receive, or share data in the system? If YES, list the system(s) and describe which data is shared.
Yes.
- Service Center Recognition/Image Processing System (SCRIPS) – SOI–DPS receives K1 image and data files from SCRIPS at the Cincinnati and Ogden hosting and editing sites. The files are received daily via EFTU.

- Modernized Tax Return Database (MTRDB) – SOI–DPS receives data files from MTRDB at the Ogden hosting and editing site. The files are pushed automatically on a weekly basis via FTP.

- Individual Master File (IMF)/ Business Master File (BMF) – SOI–DPS receives data files from Master File at the Cincinnati and Ogden hosting and editing sites. The files are pulled weekly by Team Leaders (located in Cincinnati and Ogden) via FTP.

- Large Business and International (LB&I) Data Capture System (DCS) – SOI–DPS sends images of tax returns that were scanned to the LB&I DCS application via EFTU on an as–needed basis.

- SOI also provides return document tracking information to OSPC staff for the purposes of updating the Audit Information Management System – Reference (AIMS–R) / Examination Returns Control System (ERCS) database.

11. Have the IRS systems described in Item 10 received an approved Security Certification and Privacy Impact Assessment?
Yes.

Service Center Recognition/Image Processing System (SCRIPS)
- Security Assessment & Authorization (SA&A) – June 11, 2010
- Privacy Impact Assessment (PIA) – November 23, 2009

Modernized Tax Return Database (MTRDB) (Subsystem of Modernized e–File (MeF))
- Security Assessment & Authorization (SA&A) – May 14, 2010
- Privacy Impact Assessment (PIA) – March 29, 2011

Individual Master File (IMF)
- Security Assessment & Authorization (SA&A) – March 8, 2010
- Privacy Impact Assessment (PIA) – November 10, 2009

Business Master File (BMF)
- Security Assessment & Authorization (SA&A) – June 14, 2010
- Privacy Impact Assessment (PIA) – March 16, 2010

Large Business and International (LB&I) Data Capture System (DCS)
- Security Assessment & Authorization (SA&A) – October 25, 2010
- Privacy Impact Assessment (PIA) – May 26, 2010

12. Will other agencies provide, receive, or share data in any form with this system?
Yes. Personnel from OTA have direct access to all of the microdata for the various SOI studies. Also, certain data are accessible by staff at the JCT and the BEA in accordance with legislative mandates. SOI–DPS does not receive data from Federal, State, or local Agencies.

Administrative Controls of Data

13. What are the procedures for eliminating the data at the end of the retention period?
A request for records disposition authority for SOI DPS data and associated records has been submitted (with the assistance of the IRS Records and Information Management (RIM) Program Office) to the National Archives and Records Administration (NARA), and is pending approval under registered Job No. N1–58–09–97. The SF 115 requests a permanent disposition for system data, to be transferred to NARA when 30 years old. When approved, disposition instructions for SOI DPS inputs, system data, outputs and system documentation will be published under IRM 1.15.19 Records Control Schedule for the Enterprise Computing Center – Martinsburg, Item 80.

14. Will this system use technology in a new way?
No. SOI–DPS does not use technology in a new way.

15. Will this system be used to identify or locate individuals or groups? If so, describe the business purpose for this capability.
No, this system is not used to identify or locate individuals or groups. There are data on the system that can provide the capability to identify or locate individuals or groups of people. However, this

system of records is used solely for statistical purposes and not for compliance or any other "monitoring" activity. Any data released to the public are formatted and designed to prevent the identification or location of any particular taxpayer or group of taxpayers.

16. Will this system provide the capability to monitor individuals or groups? If yes, describe the business purpose for this capability and the controls established to prevent unauthorized monitoring.
Yes. There are data on the system that can provide the capability to monitor individuals and groups. For example, certain SOI panel studies track selected taxpayers over time. However, this system of records is used solely for statistical purposes and not for compliance or any other "monitoring" activity. Any data released to the public are formatted and designed to prevent the monitoring of any particular taxpayer or group of taxpayers.

17. Can use of the system allow IRS to treat taxpayers, employees, or others, differently?
No. There is no possibility of disparate treatment of individuals or groups because the statistics, studies, and compilations are designed so as to prevent disclosure of any particular taxpayer's identity.

18. Does the system ensure "due process" by allowing affected parties to respond to any negative determination, prior to final action?
Not Applicable. The system ensures equitable treatment because the data collected are for statistical rather than tax administration purposes. Any data released to the public are formatted and designed to prevent the identification of any particular taxpayer.

19. If the system is web–based, does it use persistent cookies or other tracking devices to identify web visitors?
Not Applicable. SOI–DPS is not a public web–based system. Although some application components are web–based within SOI–DPS, none of these components allows access by any user who has not already been granted such access via either IRS Form 5081 or the SOI–DPS User Registration/Change Request Form. There is no need to track anything but pre–authorized users, as there are no "web visitors."

View other PIAs on IRS.gov