Intelligent Business Solutions Advanced Research Lab (IBS ARL) – Privacy Impact Assessment

PIA Date – Aug. 10, 2009

System Overview
The Intelligent Business Solutions Advanced Research Lab (IBS ARL) baseline environment provides a means to test analytic tools that can be used to support various types of research, econometric modeling, forecasting, and compliance studies based on the analysis of Internal Revenue Service (IRS) data. The hardware and software will exist in a highly secure and controlled network environment that is being configured to support the test and exploration of these analytic tools. IBS does not inherently include any end-user applications. IBS resembles a traditional development or test environment with the exception that it houses real data and provides secure remote user access and will require certification. This environment includes a diverse set of operating systems, server platforms, network connections, and databases in order to support the testing of various tools that may have differing platform and operating requirements. Databases are used to store sample sets of IRS data for testing purposes.

Systems of Records Notice (SORN):
- IRS 42.021--Compliance Programs and Project Files
- IRS 34.037--IRS Audit Trail and Security Records System

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

    A. Taxpayer:
- Name
- Taxpayer Identification Number (TIN)
- Tax return
- Tax return year information (1120, 1065 and K-1)
- Downloaded Securities and Exchange Commission (SEC) information concerning filings of forms verbiage for tax and ongoing operations of businesses: identify taxpayer by name, operations, corporation names, corporation officers and revenue stats that occur for each year (10K, Form 4s, Form 13s, 10Qs).

        1. The system stores data for analysis that assists the IRS in examining debt collection, taxation, and criminal and legal investigations. Data that can be stored in the system for analysis includes:
- Taxpayer name
- Social Security Number (SSN)
- TIN
- Address
- Debt amounts
- Collection status
- Garnishment data
- Savings account numbers
- Savings account balances
- Representative payee information
- Criminal and legal investigation data related to fraud type, severity level, insufficient funds amount, affiliated law enforcement agency,

legal proceedings, property seizure, financial account freeze, seized assets tracking, admissible evidence requirements
- Business EIN
- Business taxpayer address
- Assessed tax liability
- Payments, balance due amounts
- Refund checks sent
- withholdings
- Wages earned
- Interest income

B. Audit Trail Information - For each auditable event the audit trails capture, at minimum, the following information:
- Date and time that the event occurred;
- The unique identifier (e.g., user name,   Standard employee identifier (SEID), application name, etc) of the user or application initiating the event;
- Type of event (including all identification and authentication attempts (successful and unsuccessful) and commands directly initiated by the user); and
- Subject of the event (e.g., the user, file, or other resource affected); the action taken on that subject; and the outcome status (success or failure) of the event.

C. Other - Chief Counsel data:
- Subpoenas
- Summons
- Emails
- legal opinions associated with taxpayer names

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

A. IRS – files are received from business units such as data from Chief Counsel which is received on an encrypted drive.

B. Taxpayer – taxpayers do not provide data directly as this is a testing environment. Any taxpayer related information is obtained by the business unit or publicly available sites.

C. Other Federal Agencies – publicly available data from website hosted by the Security and Exchange Commission (SEC), etc. This data is obtained by Research, Analysis, and Statistics (RAS) personnel by using an isolated Digital Subscriber Line (DSL) segment unconnected to the IRS segment whereby they can search the web.

3. Is each data item required for the business purpose of the system? Explain.
Yes. Each data element is required for the business purpose of the system. Data is only collected to support business unit analysis. Once data is modeled and the analytic tool and formats are approved, the data is removed from the IBS ARL test/lab environment.

4. How will each data item be verified for accuracy, timeliness, and completeness?
The quality of the data relies on the business unit (or other data source) that presents it to IBS ARL for analytic modeling. Publicly available data is beyond the control of the system. The data itself is not important for the function of IBS ARL as it only uses the data to test tools and reporting (trending and analytics) formats for the business units.

5. Is there another source for the data? Explain how that source is or is not used.
Yes. The only other source outside the business units is publicly available information from sites such as the SEC. This data is also used to support data modeling.

6. Generally, how will data be retrieved by the user?
The users are testers within RAS and the business units. RAS IBS ARL lab environment testers log into the server and choose tools available on the server to analyze the data. Access is driven through either a Uniform Resource Locator (URL) or Commercial of the Shelf (COTS) product like Clementine. Business unit users access the data, and possibly run queries, through various analytic applications which they reach through a URL.

7. Is the data retrievable by a personal identifier such as name, SSN, or other unique identifier?
Yes. Data can be queried by a name or other identifier (including SSN) from the servers and databases using the data elements in future analytic tools.

<u>Access to the Data</u>
8. Who will have access to the data in the system (Users, Managers, System Administrators, Developers, Others)?

> **Role:** System Administrators
> **Permission:** View through viewing reports
>
> **Role:** Developers
> **Permission:** View through viewing reports
>
> **Role:** RAS developer/tester
> **Permission:** Read, write and delete access privileges to the Structured Query Language (SQL) databases
>
> **Role:** Tax Return Data Analysis Group
> **Permission:** Read, write and delete access privileges to the SQL databases
>
> **Role:** Eastport Analytics System Administrators
> **Permission:** Read, write and delete access privileges to the SQL databases
>
> **Role:** RAS Oracle Database Administrator
> **Permission:** Read, write access privileges to the SQL databases
>
> *Note: The Sun contractors cannot access data in the Oracle database.*

9. How is access to the data by a user determined and by whom?
RAS application POCs and management determines which users will be allowed access. The decision is based on need to know for the applicant. Employees need to complete an Online 5081 before they are granted access. The System Owner must approve privileged access.

Contractors undergo a National Agency Check and Inquiry Investigation plus a Credit Check (NACIC).

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.
No. Currently no other IRS systems provide, receive or share data in the system. In the future, it may be possible that Compliance Data Warehouse (CDW) will provide information.

11. Have the IRS systems described in Item 10 received an approved Security Certification and Privacy Impact Assessment?
No. Currently no other IRS systems provide, receive or share data in the system.

12. Will other agencies provide, receive, or share data in any form with this system?
Yes. IBS ARL receives information from the SEC through a portal that is segregated from the IRS network. This Information is not sent directly to the system, but may be retrieved as text files (some may also come in as Excel or HyperText Markup Language (HTML) data).

Administrative Controls of Data

13. What are the procedures for eliminating the data at the end of the retention period? IBS ARL Data is not retained in accordance with IRM retention schedules. It is only held for a short period for testing purposes and then a Guttmann algorithm is employed to remove data in preparation for the next testing session.

14. Will this system use technology in a new way?
No. IBS ARL is strictly a test and evaluation environment and will only be emulating the production environments for a given system.

15. Will this system be used to identify or locate individuals or groups? If so, describe the business purpose for this capability.
Yes. Analytics are sometimes performed to uncover potential or existing tax evasion schemes and shared with the business unit requesting the data modeling or tool testing.

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.
No. All information is static; it is not used to continually monitor an individual or business.

17. Can use of the system allow IRS to treat taxpayers, employees, or others, differently?
No. IBS ARL is strictly a test and evaluation environment and will only be emulating the production environments for a given system. The data within the system is for testing purposes only.

18. Does the system ensure "due process" by allowing affected parties to respond to any negative determination, prior to final action?
No. Not applicable because IBS ARL is only used for analytical purposes – any data used would be turned over to other business units requesting modeling and tool testing to take any actions (e.g., criminal investigations).

19. If the system is web-based, does it use persistent cookies or other tracking devices to identify web visitors?

IBS ARL has only an Intranet connection available. Uniform Resource Locators (URLs) are assigned based on the server name to allow access to other IRS Business Units as required. The isolated DSL segment to the Internet is not for public users/visitors but for use by RAS personnel to conduct outgoing web searches.

View other PIAs on IRS.gov