

**PIA Approval Date – May 26, 2010**

**System Overview:**

The LMSB Data Capture System (DCS–2) provides a tool for LMSB and other Business Operating Divisions to access data extracted from large and mid–size business tax returns that have been imaged by the Statistics of Income Distributed Processing System (SOI DPS) (LMSB Image Net (LIN) component) process. The extracted data include Schedule M–3, which was mandated by the commissioner, forms 1120, 1120S, 1118, 5471, 8886A, and 5472. Initially the data will be used to provide for much more robust modeling of compliance risks, make issue identification more rapid and accurate, and provide critical data needed by IMS, SOI, Research Personnel, and field personnel. Additionally, DCS–2 will provide a data store that will be immediately available to Modernized E–file (MeF).

**Systems of Records Notice (SORN):**

- IRS 22.026--Form 1042–S Index by Name of Treasury/Recipient
- IRS 22.027--Foreign Information System
- IRS 24.013--Combined Account Number File
- IRS 24.046--CADE Business Master File
- IRS 34.037--IRS Audit Trail and Security Records System
- IRS 36.003--General Payroll and Personnel Records (covers CADS)
- IRS 42.001--Examination Administrative Files (covers EOADS)
- IRS 42.008--Audit Information Management System (AIMS)
- IRS 42.017--International Enforcement Program Files
- IRS 42.021--Compliance Programs and Project files
- IRS 42.027--Data on Taxpayers Filing on Foreign Holdings

**Data in the System**

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

- A. Taxpayer – LMSB DCS contains corporate taxpayer data that consists of one or more tax returns for the same taxpaying entity. LMSB DCS does not collect, use, store nor share personally identifiable information (PII). Types of data available for use by LMSB DCS: Business TIN, Business Name, every tax return line item and filing period; address; relationship; published financial statements filed with the Security Exchange Commission; percent of stock ownership by corporate officers.

The tax forms used in LMSB DCS are:

- 1042 Withholding on US Source Income;
- 1065 Partnership Return and
- 1065B Partners Share K–1;
- 1120 Corporate family of returns including all forms, schedules, and attachments to the corporate filing;
- 5471 Information Return of Persons with Respect to Certain Foreign Corporation;
- 5472 Information Return of a Foreign Owned Corporation;
- 8886 Tax Shelter Disclosure Statement Form; and
- Modernized e–Filed data and traditional e–filed data.

LMSB DCS is a system to collect and analyze Corporate, Business and Industry tax data. No individual taxpayer sensitive but unclassified data is available for use by LMSB DCS.

B. Audit Trail Information –User Log–in ID and timestamp are captured in the audit trail log of the application. The audit function also provides a record of the actions performed by users in LMSB DCS and will collect data on everything viewed or accessed while on the system.

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

### A. IRS :

- Paper Returns – IRS Data is obtained from the following forms: 1042 Withholding on US Source Income; 1065 Partnership Return and 1065B Partners Share K–1; 1120 Corporate family of returns including all forms, schedules, and attachments to the corporate filing; 5471 Information Return of Persons with Respect to Certain Foreign Corporation; 5472 Information Return of a Foreign Owned Corporation; and 8886 Tax Shelter Disclosure Statement Form. Paper returns are scanned in and Optical Character Recognition (OCR) is run on the images to convert the tax information into electronic data that is stored in LMSB DCS. OCR has the capability to extract 11,000 data elements from the paper returns.
- Audit Information Management System (AIMS) (closed case records) – AIMS provides LMSB DCS with the organization code for the organization that is assigned a particular case.
- Integrated Production Model (IPM) – IPM provides current and historical data from TC590. in one consolidated read–only database for access by LMSB DCS from which LMSB will query the database to retrieve data on business return tax related information for case file selection and delivery to the field. LMSB DCS will also acquire Business Filer Model (BFM) Business Return Transaction File (BRTF) data (data transcribed from corporate tax returns) from IPM.
- MeF: The Modernized Tax Return Data Base (M–TRDB), a component of MeF, is the authoritative source of accepted returns and extensions submitted through the MeF system.
- IMS: IMS provides LMSB DCS with data on issues that were found with associated company tax return audits.

B. Taxpayer – Taxpayer company name, address, telephone number, corporate employer identification number (EIN) and other completed line items on the tax return and any attached schedules as filed by the taxpayer or his representative are maintained within the system. This includes data in both electronically–submitted and paper–filed returns from the following forms:

- 1120 Family of U.S. Corporation Income Tax Returns (F1120, S, F, FSC, PC, POL, REIT, RIC, L, ND) and 1065 monthly refresh, including MeF returns and data electronically extracted from paper filings.
- Tax Form 851 Affiliations Schedule including subsidiary returns within an MeF return
- 5471 – PSC Information Return of U.S. Persons With Respect to Certain Foreign Corporation
- 5472 – PSC Information Return of a Foreign Owned Corporation
- SOI – 1118 Foreign Tax Credit – Corporations
- SOI – 5471, 5472
- Tax Form K–1 Partners Share of Partnership or Subchapter S income
- 8886 Tax Shelter Disclosure Statement Form

### C. Other Third Party Sources:

- North American Industry Classification System (NAICS) table: NAICS is a uniform numbering system to which all businesses can subscribe. This system allows others to know under what industry the businesses fall. An excel spreadsheet is downloaded from NAICS, scanned, and loaded into LMSB DCS annually.
- Standard & Poor's: Public financial data is purchased from Standard & Poor's and is acquired via a portable hard drive which is encrypted with WinZip 9 Federal Information Processing Standards (FIPS) approved encryption. This includes the business' Securities and Exchange Commission (SEC) EDGAR filings data.

### 3. Is each data item required for the business purpose of the system? Explain.

Yes. The LMSB DCS will provide LMSB personnel with systems that capture and provide information on business tax return issues to examination personnel, managers, and research analysts. The issue-based management strategy will also support IRS strategic goals of reducing taxpayer burden, increasing the productivity of examination personnel and aid in the recruitment and retention of a skilled and more satisfied workforce. Corporations, Businesses and Industries will not be required to retrieve and provide paper SEC filings and documents, thereby, reducing their compliance burden. Agents will not have to manually input data from the paper documents thereby reducing the time it takes to perform their duties.

LMSB DCS consists of:

- An SQL 2005/2008 database of corporate, business and industry tax return, financial and interrelationship data linked to SQL 2005/2008 Reporting and Analysis Services server that provide ad hoc reporting capability to LMSB Employees.
- SPSS (Predictive Analytics Software), SAS (Business Analytics Software), Clementine, and ProClarity Statistical software

LMSB DCS will support workload and issue management. It will also support research on emerging global and tax shelter issues and predictive models for unreported income, compliance risks, and the effectiveness of pre-filing initiatives. LMSB DCS data is transferred to the IRS Selectable Workload Classification (SWC) System via EFTU for use in case classification.

### 4. How will each data item be verified for accuracy, timeliness, and completeness?

The LMSB DCS system utilizes copies of "authoritative data" from IRS data sources. Data is downloaded and run through the system. It is supplemented by data extracted from paper returns using an OCR process with multiple rules engines, a system that executes business rules, to ensure accuracy. LMSB DCS data verification occurs in four phases: OCR Verification; Manual Validation; Managerial Review; and Data Analysis.

Two types of validation errors are identified during OCR processing and result in an additional indicator to a verifier:

- Fields where a character or symbol does not meet a percentage of recognition points. For example, the program may only be 50% sure that a number is a 3 as opposed to an 8 (70% is the default).
- A field does not meet predefined business rules. For example, the EIN on the return must match the EIN on the return coversheet.

During the Verification: During the Manual Validation phase, the verifier must stop on each field and press enter to move forward. Fields identified as problematic during OCR are flagged to the verifier. Recognition problems will not allow the verifier to enter past the field until corrected. Field values that violate business rules cause an error message to appear. Verifier corrections are further evaluated

against the business rules in the background. The error must be corrected or the message bypassed (appropriate when the taxpayer made the error – i.e. we do not alter returns).

The next phase is the Managerial Review. After the verification stage, a sample of returns (normally 10%) is sent to a review queue for managers or their representatives. The manager can compare the return images to what the verifier transcribed. Errors found are referred to the data development team for analysis and corrective action. Sample rates can be changed per workflow and specific verifier work can be sent to the review queue in order to identify transcription issues quickly.

The final phase is the Data Analysis Phase. During this phase, exported data is analyzed for the following types of errors:

- Errors not readily handled by business rules. For example, correcting missing unique key fields used to link records together.
- Error conditions inappropriately bypassed or missed by verifiers.

When errors are found, one of the following actions may be taken:

- Recycle the batch to verification. This usually occurs when more than one transcription error is found on a single page of the return. Analysis of data usually bears out a less than 1% chance of error after verification. If more than one error occurs, this decreases confidence in the validity of the rest of the return. We notify the verifier's manager if we see a pattern of returns with high error rates getting past the same verifier.
- Make small corrections to the exported data. This will occur if there is an error in only one field but there is confidence that the rest of the return is valid.
- Make no correction. This usually occurs when we see that a business rule was bypassed but is due to taxpayer reporting and not transcription error.

Data is manually inputted into staging servers by verifiers and checked prior to going live into the database. Once data enters the staging server it cannot be modified and becomes "read-only" data. Users access the data via an SQL management tool that permits them to view data based on work assignments.

**5. Is there another source for the data? Explain how that source is or is not used.**

There are no other sources for the data.

**6. Generally, how will data be retrieved by the user?**

Data can be retrieved by the corporate, business or industry EIN number, business address, NAICS code, or stock ticker symbol, or a transactional relationship to an entity under examination, depending upon authorized user access permissions.

**7. Is the data retrievable by a personal identifier such as name, SSN, or other unique identifier?**

Yes. Taxpayer Identification Number (TIN/EIN) is used for retrieval of data along with industry type, geographic location, and Stock ticker symbol or Standard Industrial Classification (SIC) code. No individual Social Security numbers (SSN) or other identifiers are collected, used, stored, or shared.

**Access to the Data**

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

Only employees having the need to know and the right to know will have access to the data in the system. The following are DCS Users:

**Role:** Enterprise Operations (EOps) System Administrators  
**Permission:** System Administrators have access to hardware.

**Role:** EOps Database Administrators  
**Permission:** Database Administrators have access to feeds and exports for maintaining data.

**Role:** LMSB DCS Users  
**Permission:** LMSB DCS users have access to OCR data to verify the data.  
Researcher/statisticians may also have access to certain databases depending on what access their management grants them.

**Role:** LMSB DCS Install  
**Permission:** LMSB DCS Install users have installation permissions to update software.

**Role:** LMSB DCS Classify  
**Permission:** LMSB DCS Classify users have access to manually classify returns and verify the information is accurate.

**Role:** LMSB DCS Software Admin  
**Permission:** LMSB DCS Software Administrators have access for uploading extracts received. Additionally, since tax forms change slightly every year, these users make updates to the templates so that OCR can pull data from paper returns accurately.

**Role:** LMSB Strategic Research Program Planning (SRPP) DataMart Admin  
**Permission:** LMSB SRPP DataMart Administrators perform SQL Server database administration functions including configuration, database creation and backup, and loading of data.

**Role:** LMSB SRPP DataMart Developers  
**Permission:** LMSB SRPP DataMart Developers have read/write permission to own schema in workgroup database. Each developer have own schema to develop tables and load information into those tables. Own schema prevents others from viewing their data.

**Role:** LMSB SRPP DataMart Users  
**Permission:** LMSB SRPP DataMart users have read permission to RWI datamart, which includes research and workload identification data.

**Role:** LMSB SRPP APP Admin  
**Permission:** LMSB SRPP APP Administrators perform administration functions for application services in Apps Server.

**Role:** LMSB SRPP APP Users  
**Permission:** LMSB SRPP APP users have access to and can use applications hosted in Apps Server.

## **9. How is access to the data by a user determined and by whom?**

All access credential requests are enforced through the Online 5081 process for granting permissions to systems and applications used by IRS personnel. A formal request is made through the IRS employees' management chain. Online 5081 forms are completed. Each request is evaluated and a determination to grant access or deny access is made. IRS employees who are authorized on the system will have access to the data. LMSB DCS Administrators utilize the role based access features

on the COTS software product to limit what users can retrieve. Users are only permitted to access data for EINs authorized by their manager.

**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.

- Issue Management System (IMS) –The Issue Management System (IMS) and Issue Based Management Information System (IBMIS), a component of IMS, provide LMSB DCS with data on issues that were found with associated company tax return audits.
- Integrated Production Model (IPM) – LMSB DCS database will access the IPM database using an established database link. IPM provides current and historical data from TC590 in one consolidated read-only database for access by LMSB DCS from which LMSB will query the database to retrieve data on business return tax related information for case file selection and delivery to the field. LMSB DCS will also acquire Business Filer Model (BFM) Business Return Transaction File (BRTF) data (data transcribed from corporate tax returns) from IPM.
- Modernized e-File (MeF) – The Modernized Tax Return Data Base (M-TRDB), a component of MeF, is the authoritative store of accepted returns and extensions submitted through the MeF system. MeF data is received from an Enterprise File Transfer utility (EFTU) data transfer from the M-TRDB system to the Ogden Service Center where LMSB DCS acquires them after data perfection software routines are run against the data.
- Audit Information Management System (AIMS) – AIMS is a closed case records application that provides LMSB DCS with the organization code for the organization assigned to a particular case.
- Selectable Workload Classification (SWC) – SWC receives data from LMSB DCS via an EFTU transfer. SWC runs business rules against the data to select businesses for audit.

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

Yes.

**Audit Information Management System Reference (AIMS-R)**

- Certification & Accreditation (C&A) – May 1, 2009
- Privacy Impact Assessment (PIA) – February 11, 2009

**Issue Management System (IMS)**

- Certification & Accreditation (C&A) – August 8, 2007
- Privacy Impact Assessment (PIA) – June 1, 2008

**Integrated Production Model (IPM)**

- Certification & Accreditation (C&A) – August 4, 2008
- Privacy Impact Assessment (PIA) – September 12, 2008

**Modernized e-File (MeF)**

- Certification & Accreditation (C&A) – May 9, 2007
- Privacy Impact Assessment (PIA) – October 23, 2009

#### Selectable Workload Classification (SWC)

- Certification & Accreditation (C&A) for SWC Component 1 completed on November 20, 2009 and Certification & Accreditation (C&A) for SWC Component 2 completed on February 2, 2009
- Privacy Impact Assessment (PIA) –TBD

#### **12. Will other agencies provide, receive, or share data in any form with this system?**

No, other agencies will not provide, receive, nor share data in any form with this system.

#### **Administrative Controls of Data**

#### **13. What are the procedures for eliminating the data at the end of the retention period?**

The system is defined as a recordkeeping system under 36 CFR Chapter XII. All records housed in the system will be erased or purged from the system at the conclusion of their retention period(s) as required under IRM 1.15.6. Under design at the time of this publication, individual (Form 1042, and Partnership Return – Form 1065) tax filing information will be eliminated from the system in accordance with IRS Records Control Schedule 1.15.29, Item 55 for Electronically Filed Individual Partnership and Fiduciary Income Tax Returns. Corporate (Form 1120) tax filing information will be eliminated from the system using NARA's recently approved IRS Records Control Schedule IRM 1.15.19, Item 81 for Modernized e-File System (Job No. N1-58-09-98, approved by NARA 5/13/10 – not yet published in IRM). IRS Records Control Schedule 1.15.29, Item 344 will be followed for the elimination of Information Returns for Foreign Corporation, and Examinations and Audits Case Files will be eliminated in accordance with IRS Records Control Schedule 1.15.23, Item 42. A Standard Form (SF) 115 Request for Records Disposition Authority for Issue Management System records is currently under review at the National Archives (registered as Job No. N1-58-09-105), with approval expected in the near term. Other records series added to the system in future updates to the program will be managed according to requirements under IRM 1.15.1 and 1.15.6 and will be destroyed using IRS Records Control Schedules 1.15.8 through 1.15.62 as coordinated with the IRS Records and Information Management Program and IRS Records Officer.

#### **14. Will this system use technology in a new way?**

No, the system will 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, the system will not be used to identify nor locate individuals or groups.

#### **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, the system does not provide the capability to monitor individuals or groups.

#### **17. Can use of the system allow IRS to treat taxpayers, employees, or others, differently?**

No, the system does not allow IRS to treat taxpayers, employees or others differently.

#### **18. Does the system ensure "due process" by allowing affected parties to respond to any negative determination, prior to final action?**

Yes. Corporations, businesses, and Industries are afforded legal rights in the same way as due process protects individuals.

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

The system is Intranet Web-services based but does not use “cookies” or other tracking devices. Items viewed are logged in the security files. Internal IP addresses are not viewable outside the IRS.

[View other PIAs on IRS.gov](#)