U.S. Securities and Exchange Commission

Enterprise Data Warehouse (EDW) Platform
PRIVACY IMPACT ASSESSMENT (PIA)

May 31, 2020

Office of Information Technology

1. Name of Project or System.
   Enterprise Data Warehouse (EDW) Platform

2. Describe the project <u>and</u> its purpose or function in the SEC's IT environment.
   The Enterprise Data Warehouse (EDW) infrastructure will enable the provisioning of data to Commission staff for search and analysis through a virtual data warehouse platform. As part of its Enterprise Data Warehouse initiative, the SEC will begin to integrate data from various systems to provide more comprehensive management and financial reporting on a regular basis and facilitate better decision-making. (SEC Strategic Plan, 2014-2018, pg.31) The EDW will replicate source system-provided data from other operational SEC systems and provide a simplified way of producing Agency reports.

   The EDW is designed for summarization of information, retrieval of information, reporting speed, and ease of use (i.e. creating reports). It will enhance business intelligence, augment data quality and consistency, and generate a high return on investment by allowing users to quickly search and access critical data from a single location and obtain historical intelligence to analyze different time periods and performance trends in order to make future predictions. (SEC FY2013 Agency Financial Report, pg.31-32)

   The data content in the Enterprise Data Warehouse platform will be provisioned incrementally over time via a series of distinct development projects, each of which will target specific data sets based on requirements provided by SEC business stakeholders.

   The data content for these projects will vary in nature including publicly available data, highly sensitive Mission data, and personally identifiable information (PII).   It is therefore expected that each new project on the Enterprise Data Warehouse platform will be reviewed from a security and privacy perspective based on the projects unique data requirements.

3. Requested Operational Date? 6/10/2013. This PIA documents the privacy impact associated with data in the EDW since its inception and future data provisioning projects.

4. System of Records Notice (SORN) number? Since EDW does not collect information directly from individuals and no new information is created through the system about individuals, any personally identifiable information (PII) in the EDW will be collected from an existing SEC system and will be covered by a SORN listed below:

   - SEC-01 "Registration Statements Filed Pursuant to Provisions of the Securities Act of 1933, Securities Exchange Act of 1934, Public Utility Holding Company Act of 1935, and Investment Company Act of 1940"
   - SEC-02 "Applications for Registration or Exemption under the Investment Company Act of 1940"
   - SEC-03 "Notification of Exemption from Registration under the Securities Act of 1933"

- SEC-04 "Beneficial Ownership, Acquisition, Tender Offer, and Solicitation Records Filed under the Securities Exchange Act of 1934"
- SEC-05 "Ownership Reports and Insider Trading Transactions Records Filed under the Securities Exchange Act of 1934, Public Utility Holding Company Act of 1935, and Investment Company Act of 1940"
- SEC-06 "Periodic Reports Filed under the Securities Act of 1933, Securities Exchange Act of 1934, Public Utility Holding Company Act of 1935, and Investment Company Act of 1940 and Investment Advisors Act of 1940"
- SEC-07 "Proposed Sale of Securities Notices Filed under the Securities Act of 1933"
- SEC-08 "Proxy Soliciting Material Filed under the Securities Exchange Act of 1934, Public Utility Holding Company Act of 1935, and Investment Company Act of 1940"
- SEC-09 "Correspondence Files Pertaining to Registered Broker-Dealers"
- SEC-10 "Correspondence Files Pertaining to Registered Investment Advisers"
- SEC-11 "Correspondence Files Pertaining to Registered Investment Companies"
- SEC-15 "Payroll, Attendance, Retirement, and Leave Records"
- SEC-20 "Division of Corporation Finance Index for Filings on Schedule 13D and Filings under Regulations A and B"
- SEC-39 "Personnel Management Employment and Staffing Files"
- SEC-40 "Personnel Management Training Files"
- SEC-42 "Enforcement Files"
- SEC-49 "Broker-Dealer Records"
- SEC-50 "Investment Adviser Records"
- SEC-55 "Information Pertaining or Relevant to SEC Registrants and Their Activities"
- SEC-56 "Mailing, Contact and Other Lists"
- SEC-61 "Municipal Advisor Records"
- SEC-62 "Correspondence Files Pertaining to Municipal Advisors; Municipal Advisor Logs"
- SEC-63 "Tips, Complaints, and Referrals (TCR) Records"
- GSA/GOVT- 7 "Personal Identity Verification Identity Management System (PIV IDMS)"
- OPM/GOVT-1 "General Personnel Records"
- OPM/GOVT-5 "Recruiting, Examining, and Placement Records"

5. Is this an Exhibit 300 project or system? ☒ No ☐ Yes

6. What specific legal authorities, arrangements, and/or agreements allow the collection of this information? The underlying SORNs for the various data sets, if applicable, provide the legal authority for the collection of the data.

Specific Questions

SECTION I - Data in the System
1. What data about individuals could be collected, generated, or retained?
   The EDW system collects, uses, maintains, retrieves, and disseminates the following information:
   Tracking and Reporting Examination National Documentation System (TRENDS)

- Documents relating to examinations
- Individuals bank and brokerage account statements
- Name
- Address
- Telephone number

Tips, Complaints, and Referrals (TCR)
- Individuals name
- Date of birth
- Social security number
- Mailing address
- Telephone number
- Email address
- Tip, complaint and referral information - including allegation description, dates, and supporting details related to individuals
- Web forms
- Criminal history
- Working papers of the staff and other documents and records relating to the matter

The HUB System
- Data collected pertains to matters under inquiry, investigations, and actions
- Internal identification and status information (i.e. matter name, matter number, date opened and closed)
- Staffing information
- Narrative summaries
- Substantive categorizations of the matter (i.e. Insider Trading)
- Procedural steps taken (i.e. access grants, investigative testimony, Wells notices)
- Links to key documents (i.e. Complaints, Orders Instituting)
- Court and filing information for actions, including (i.e. forum, date filed, violations alleged, relief sought, and outcome of actions)
- Social security numbers

EDGAR EDR Entity Tables
- Name
- Date of birth
- Mailing address
- Telephone number
- Email address
- Social security number
- Bank account number
- Financial information
- Password
- Certificate/license number
- Educational information
- Fax number
- Tax identification number
- Employment history

- Disciplinary history

EDGAR MOMENTUM Reporting Tables:
- General Ledger Entries
- Vendor accounts
- Vendor address
- SEC account activities
- SEC General Ledger Postings
- Accounting transaction details
- Valid document types
- Billing documents
- Fund and payments
- Cash Receipt documents

Security Master DB - (PII related to):
- Investment advisors
- Brokers and dealers
- Transfer agents
- Investment companies
- Publicly traded companies
- Municipal securities
- Dealers,
- Government security
- Clearing agencies
- Exchange
- Securities information processor
- Nationally recognized statistical rating agencies

DERA FINRA Datamart - (PII related to):
- FINRA broker and dealer firm
- Investment advisor firm
- Individual and Branch information

FINRA PII Data Related to IA and BD:
- Individual and Branch information
- Social Security Number
- Date of Birth

FINRA Bluesheet - (PII related to):
- Name
- Mailing address
- Telephone number
- Bank account number
- Financial information
- PII that may be obtained as part of a criminal investigation
- Surveillance evidence required to determine responsibility for a crime

- Monitoring and questioning of affected parties
- Legal investigation information associated with gathering information about a given party
- Legal prosecution and litigation activities involved with presenting a case in a legal proceeding in an attempt to prove guilt/responsibility

Active Directory (SEC worker information)

- Full name
- Address (office and personal)
- Department
- Telephone number (mobile and office)
- Company
- SIP address
- SMTP address
- User account control
- Email address
- SAM Account name
- Division
- Country
- Job title
- Manager
- Home directory
- Home drive
- Primary group ID
- Last log on

DOI Datamart - WTTS

- WTTS ID
- WTTS status
- Slot number
- Unique Person Number (UPN)
- Full Name
- Social security number
- Date of birth
- Education level
- Sex
- Home address
- Department
- Org code
- Duty station code
- Series
- Pay plan
- Grade 1
- Step
- Full Performance level
- Required to fill out financial disclosure form

- Key personnel indicator
- Position number
- Position sensitivity
- OPM position title
- Security information
- Total salary
- Date of security clearance
- Date of ethics clearance
- Announcement type

DOI Datamart - FPPS
- Series
- Full name (current and previous)
- Date accession
- Work schedule
- Date (i.e. entered into current position, entered on duty bureau, entered present grade, probation)
- Bargaining unit status/description
- Date of birth
- Date of security clearance and if applicable security reinvestigation
- Date of security status
- Department
- Education level
- Employee status history
- Financial statements
- Full performance level grades
- Functional classification
- Grade or level
- Occupational series/description
- OPM pay status
- Sex
- Pay plan
- Salary and pay rate
- Position title
- Position ID
- Social security number
- Work schedule

Leading, Learning, and Performing (LEAP)
- User ID
- Active user
- Full name
- Complete date
- Item ID
- Curriculum ID
- Scheduled offering ID

- Grade
- Tuition
- Employee type ID
- Address
- Resume
- Phone number
- Social security number
- Date of birth
- Race, nationality, origin
- Education

USAccess

- Name/ aliases
- Date of birth
- Person ID
- Social security number
- US citizen (yes or no)
- Tax ID
- Foreign ID
- Birth city, state and country
- Address
- Telephone number (cell and home)
- Email address
- Gender
- Race
- Eye color
- Hair color
- Height
- Weight
- Sponsorship ID
- PIV card serial number
- PIV card personal identification number (PIN)
- Cardholder unique identification number (CHUID)
- Rank
- Employment status
- Contract number
- Employee ID
- Fingerprints
- Digital color photograph
- Country and city of birth

Financial Data Mart - (PII related to):

- Delphi (financial module)
- Prism (purchase)
- Momentum (collection fee)
- Federal travel information

PTCS Securities Data - (PII related to):
- Securities reference data used by PTCS pre-trading evaluation

OCIE FINRA IA (Investment Advisory)
- FINRA IA Firm information

Automated Bluesheet Analysis Project (ABAP) Bluesheet Data
- Name
- Mailing address
- Telephone number
- Bank account number
- Financial information
- PII that may be obtained as part of a criminal investigation
- Surveillance evidence required to determine responsibility for a crime
- Monitoring and questioning of affected parties
- Legal investigation information associated with gathering information about a given party
- Legal prosecution and litigation activities involved with presenting a case in a legal proceeding in an attempt to prove guilt/responsibility

DERA Option Data- (PII related to):
- Large option data from the OPRA (Options Price Reporting Authority)(500TB)

DERA FOCUS Data – (PII related to)
- Financial and Operational Combined Uniform Single Reports

DERA TRACE Data –(PII related to)
- Trade Reporting and Compliance Engine

DERA MF Holdings Data (PII related to)
- On demand data feed from Morningstar for analysis and reporting on holdings from various investors.

DERA CDS Data –(PII related to)
- Credit default swap data from the Trade Information Warehouse (TIW)

DERA HEDGEFUND –(PII related to)
- IA Hedge Fund

DERA N-MFP Data –(PII related to)
- Monthly schedule of portfolio holdings of money market funds report

DERA TAQ Data –(PII related to)
- Trade and Quotes data

DERA FORM D Data –(PII related to)
- Notice of Exempt Offering of Securities

2. Does the project/system use or collect the social security number (SSN)? (This includes truncated SSNs)

    ☐ No.

    ☒ Yes. If yes, provide the function of the SSN and the legal authority to collect.

    The data content for Enterprise Data Warehouse projects will vary in nature based on each project's unique business requirements. As such, personally identifiable information, such as SSN could be included in the Enterprise Data Warehouse in order to enhance the SEC's

Mission activities to enforce securities laws and regulations. The underlying SORNs for the various data sets provide the legal authority for the collection of the data.

3. What are the sources of the data?
    EDW does not collect information directly from individuals. Information collected and stored in EDW is gathered from various SEC source systems. Therefore, the sources of data will vary based on each project's unique business requirements. Sources of the data currently provisioned for the EDW are listed in Section I, Question 1.

    Please see appendix for a current listing of data sources and data elements.

4. Why is the data being collected?
    EDW collects information from specific data marts to provide reports that source systems' owners may need to fulfill the SEC mission. The Enterprise Data Warehouse will enhance business intelligence, augment data quality and consistency, and generate a high return on investment by allowing users to quickly search and access critical data from a single location. It will also assist in obtaining historical intelligence to analyze different time periods and performance trends in order to make future predictions.

5. What technologies will be used to collect the data?
    The Enterprise Data Warehouse platform is an integrated technology solution which is composed of the following tools:

    •IBM InfoSphere Information Server (IIS – version 9.1) is a data integration platform that supports understanding, cleansing, transforming and delivering trusted information (or the Extract, Transform, and Load – ETL – process to get data) to the EDW.

    •Netezza (1000-3 series; 1000-6; C1000-16; Host Platform Configuration Version 4.15.1) is a data warehouse appliance which architecturally integrates database, server and storage components into a single unit. It uses a proprietary Asymmetric Massively Parallel Processing (AMPP) architecture that combines open, blade-based servers and disk storage with a proprietary data filtering process using field-programmable gate arrays (FPGAs). This will act as the backbone for the EDW to enable a high-performance data warehouse and advanced data analytics platform.
    •IBM InfoSphere Guardium (version 8.2) provides real-time data security by assuring the privacy and integrity of the information stored in the EDW and it automates the entire compliance auditing process in heterogeneous environments.

    •IBM InfoSphere BigInsights (version 2.0.0.0) brings the power of Hadoop to the enterprise. Apache Hadoop is the open source software framework, used to reliably managing large volumes of structured and unstructured data.

SECTION II - Attributes of the Data (use and accuracy)
1. Describe the uses of the data.

Broadly, the data in the Enterprise Data Warehouse will be used to enhance business intelligence, augment data quality and consistency, and generate a high return on investment by allowing users to quickly search and access critical data from a single location and obtain historical intelligence to analyze different time periods and performance trends in order to make future predictions.

The specific uses of data provisioned to the Enterprise Data Warehouse will vary, and will be based on the unique objectives and business requirements that will drive each data provisioning project.

2. Does the system analyze data to assist users in identifying previously unknown areas of note, concern or pattern? ☐ No ☒ Yes If yes, please explain: Yes, the Enterprise Data Warehouse is intended to integrate data from disparate systems in order to identify patterns and trends which are currently difficult or impossible to discern via manual processes.

3. How will the data collected from individuals or derived by the system be checked for accuracy?
The requirements for data quality (accuracy, completeness, timeliness, validity, precision) are the responsibility of the underlying source system. These requirements will vary based on the unique business requirements of each project.

Since the data in the Enterprise Data Warehouse will be extracted from other sources, the quality of the underlying data will depend on the quality controls present in those source systems.

SECTION III - Sharing Practices
1. Will the data be shared with any internal organizations?
☐ No ☒ Yes If yes, please list organization(s):
Yes. Broadly, the data in the Enterprise Data Warehouse is expected to be a shared enterprise asset to facilitate information sharing and knowledge management.

However, each data provisioning project will have unique business requirements around the access controls for its component data. These controls will drive which individuals within the enterprise can see particular data sets or data elements. The access controls will be role-based, and will be defined by the designated data business owner, who will also approve the on-going membership of the associated access roles through provisioning, de-provisioning, and routine access control audits.

2. Will the data be shared with any external organizations?
☒ No ☐ Yes If yes, please list organizations(s):   How is the data transmitted or disclosed to external organization(s)? EDW does not share information with external organizations. If there are reports generated by EDW shared outside of the SEC, they are shared pursuant to the source system's published routine uses within the relevant SORN.

3. How is the shared data secured by external recipients?
N/A

4. Does the project/system process or access PII in any other SEC system?
   ☐ No
   ☒ Yes. If yes, list system(s).EDW will receive PII from various SEC source systems including: TCR, TRENDS, EDGAR EDR, Security Master Entity DB, ABAP Bluesheet data, FINRA DataMart, various DERA data sets, Employee data, general ledger entries and back-end financials. EDW does not share PII with any other SEC systems. The data content for Enterprise Data Warehouse projects will vary in nature based on each project's unique business requirements.

   Please see appendix for a current list of EDW data sources and data inventory.

SECTION IV - Notice to Individuals to Decline/Consent Use

1. What privacy notice was provided to the different individuals prior to collection of data? (Check all that apply)
   ☐ Privacy Act Statement ☐ System of Records Notice ☒ Privacy Impact Assessment
   ☐ Web Privacy Policy ☐ Notice was not provided to individuals prior to collection

2. Do individuals have the opportunity and/or right to decline to provide data?
   ☐ Yes ☐ No ☒ N/A
   Please explain: EDW only provides reporting technology. Any opportunities for individuals to decline to provide information or opt out of EDW reporting would be processed through the individual source system.

3. Do individuals have the right to consent to particular uses of the data?
   ☐ Yes ☐ No ☒ N/A
   Please explain: EDW only provides reporting technology. Any opportunities for individuals to consent to uses would be processed through the individual source system.

SECTION V - Access to Data (administrative and technological controls)

1. Has the retention schedule been established by the National Archives and Records Administration (NARA)?
   ☐ No If no, please explain:
   ☒ Yes If yes, list retention period: The data retention periods for Enterprise Data Warehouse projects will vary in nature based on each project's unique business requirements.

   Each source system and its responsible data business owner will define its data retention requirements in partnership with the Office of Records Management. As data is added or deleted in the source systems, it is also added or deleted in EDW during regular updates and refreshes. Data is retained within EDW as long as indicated by each source system's NARA-approved records retention and disposal schedule.

2. What are the procedures for identification and disposition of the data at the end of the retention period?
   Data will be removed from the Enterprise Data Warehouse following the end of the retention period.

3. Describe the privacy training provided to users, either generally or specifically relevant to the program or system?
   Training is provided by the Privacy Office which all SEC employees and contractors must complete yearly. The training outlines their roles and responsibilities for properly handling and protecting PII. The user roles governing the access to sensitive or PII data in the Enterprise Data Warehouse will be outlined within the standard operating procedure (SOP) document of each data provisioning project.

4. Has a system security plan been completed for the information system(s) supporting the project?
   ☒ Yes If yes, please provide date SA&A was completed: 6/10/2014
   ☐ No If the project does not trigger the SA&A requirement, state that along with an explanation

5. Is the system exposed to the Internet without going through VPN?
   ☒ No ☐ Yes If yes, Is secure authentication required? ☐No ☐Yes; and
                        Is the session encrypted? ☐ No ☐Yes

6. Are there regular (ie. periodic, recurring, etc.) PII data extractions from the system?
   ☒ No   ☐ Yes If yes, please explain:

7. Which user group(s) will have access to the system?
   The specific users of the Enterprise Data Warehouse will be identified through each data provisioning project based on their unique business requirements and there will be access controls defined for its component data. These controls will drive which individuals within the enterprise can see particular data sets or data elements. The access controls will be role-based, and will be defined by the designated data business owner, who will also approve the on-going membership of the associated access roles through provisioning, de-provisioning, and routine access control audits.

8. How is access to the data by a user determined?
   The access controls for each project will be role-based, and will be defined by the designated data business owner, who will also approve the on-going membership of the associated access roles through provisioning, de-provisioning, and routine access control audits. The user roles governing the access to sensitive or PII data in the Enterprise Data Warehouse will be outlined within the SOP document of each data provisioning project.

   Are procedures documented? ☒ Yes ☐ No
9. How are the actual assignments of roles and rules verified.
   Role assignments will be determined by the business data owner as part of the requirements definition process and validated through testing as part of the SDLC process. On-going validation of roles will be conducted via regular audits.

10. What auditing measures/controls and technical safeguards are in place to prevent misuse (e.g., unauthorized browsing) of data?

The access controls will be role-based, and will be defined by the designated data business owner, who will also approve the on-going membership of the associated access roles through provisioning, de-provisioning, and routine access control audits on an annual basis.

SECTION VI - Privacy Analysis

Given the amount and type of data being collected, discuss what privacy risks were identified and how they were mitigated.

The data content for Enterprise Data Warehouse projects will vary in nature, and may include publicly available data, highly sensitive Mission data, and/or personally identifiable information (PII).   It is therefore expected that each new project on the Enterprise Data Warehouse platform will be reviewed from a security and privacy perspective based on the projects unique data requirements.

If sensitive or PII data is required as defined by the business data owner, access controls will be defined around the data. These controls will dictate which individuals within the enterprise can see particular data sets or data elements. The access controls will be role-based, and will be defined by the designated data business owner, who will also approve the on-going membership of the associated access roles through provisioning, de-provisioning, and routine access control audits.

There is also a privacy risk that EDW may receive more information than is needed to provide reports requested or required by the user access and interface application. This risk is mitigated by EDW only receiving the relevant data elements from the source systems. Information is separated by data marts accessible only by users approved by the source systems' data steward with an established need to know. This process limits users from creating reports that contain more information than is needed for the report supporting their specific mission. The source systems collect information on individuals pursuant to the source systems guidelines and legislative limitations. No additional information is collected.