

U.S. Securities and Exchange Commission

Palantir Enterprise Data Analytics Platform (EDAP)

PRIVACY IMPACT ASSESSMENT (PIA)



March 18, 2014

Privacy Impact Assessment
Palantir - Enterprise Data Analytics Platform (EDAP)

General Information

1. Name of Project or System.
Palantir Enterprise Data Analytics Platform (EDAP)
2. Describe the project and its purpose or function in the SEC's IT environment.
The SEC Enterprise Data Analytics Platform (EDAP) is an enterprise wide single-platform analytic software tool that provides SEC the capability of integrating structured, unstructured, and semi-structured data from a wide variety of data sources for seamless research and analysis in one unified environment. SEC investigators can use the analytical platform to find, analyze, and visualize connections between disparate sets of data to uncover suspicious behavior in securities-related activities and quickly trace its origin. This is accomplished through enterprise-wide data integration, advanced search and discovery, and secure collaboration across the Commission.

The purpose of EDAP is to support investigations into securities and related fraud, to support inspection of regulated entities, to inform SEC policy development related to financial sector oversight, to support the presentation in a court of law of analysis conducted during investigations, and to review reporting requirements to ensure sufficient information is collected to support analysis. Initially, the focus will be in five areas of analysis: (i) Tips, Complaints, and Referrals, (ii) Trading Data (ABAP), (iii) Accounting Fraud Analysis, (iv) Foreign Corrupt Practices Act (FCPA) Fraud Analysis, and (v) analysis of EDGAR (filing) data.

The goal is to provide the SEC with robust data integration, search and analysis, knowledge management, secure collaboration, and algorithmic engine capabilities. These enhanced capabilities should lead to improved compliance activities (through better examinations) and enforcement actions (through better targeting and investigation of alleged wrongdoing).

3. Requested Operational Date? EDAP currently is operating in a limited capacity under a waiver that is due to expire in January. An authorization to test (ATT) for move into production was granted on December 19, 2013 and the targeted Authorization to Operate (ATO) date is February 15, 2014. Meanwhile, an extension of the waiver to April 2014 is currently being sought. Enterprise roll-out will not occur until the ATO is obtained.
4. System of Records Notice (SORN) number? EDAP will not be the original source of data. Data will be collected in accordance with the respective, governing System of Records Notice (SORN) of the source system from which the data is derived.
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? 15 U.S.C. 77a et seq., 78a et seq., 80a-1 et seq., 80b-1 et seq., and 5 U.S.C. 302.

Specific Questions

SECTION I - Data in the System

1. What data about individuals could be collected, generated, or retained?

Privacy Impact Assessment

Palantir - Enterprise Data Analytics Platform (EDAP)

All underlying data in EDAP comes from other SEC systems. Therefore, EDAP is a pass-through vehicle used for Enterprise analytical and investigative purposes. Underlying PII in 9 SEC systems that EDAP will pull from include but are not limited to the following: individual names; dates of birth; social security numbers; addresses; telephone numbers; tip, complaint, and referral information including allegation descriptions, dates, and supporting details; supporting documentation; web forms; e-mails; criminal history; working papers of the staff; and other documents and records relating to the matter. Work and home addresses, place of employment, trading records, bank account records.

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.

Social Security Numbers are collected pursuant to Section 21(a) of the Exchange Act and related rules authorizing the Commission to conduct investigations of potential violations of the federal securities laws, to identify such persons accurately and assist in determining involvement in other matters. EDAP will utilize the SSN to cross-match individuals across multiple data sources.

3. What are the sources of the data?

EDAP includes data from IRIS, TRENDS, FINRA Datamart, Recommind, HUB, TCR, EDGAR, and ABAP/EDW, as well as case-specific data loaded by investigators. Data may be public (as in the case of much EDGAR data), obtained from a regulated entity such as an exchange or SRO (as in the case of some ABAP data), generated by the SEC (as in the case of HUB data and some TCR data), sent from the public voluntarily (as in the case of a tip or complaint) or produced in response to an administrative subpoena.

4. Why is the data being collected?

No new data is collected. The data is being compiled for use by authorized SEC personnel in receiving, recording, assigning, tracking, analyzing, and taking action on tips, complaints, and referrals, foreign corruption and bribery, stock market abuse, accounting fraud, and public company fraud received from individuals and entities related to actual or potential violations of the federal securities laws; investor harm; or conduct of public companies, securities professionals, regulated entities and associated persons. Currently much of this compilation is a manual process conducted on a case-by-case basis. Thus, it serves great purpose to SEC.

5. What technologies will be used to collect the data?

The Palantir platform does not collect data, it processes existing SEC data. Individual SEC systems listed in question 3 above collect the data. For the majority of systems, Palantir will collect the data by directly connecting to the source system's database and running a series of database queries. This will be done using Palantir Kea data integration scripts, which utilize Java's JDBC connectivity. Kea is itself a Groovy-based domain specific language for integrating data into Palantir. In the case of Recommind, which does not have an accessible

Privacy Impact Assessment

Palantir - Enterprise Data Analytics Platform (EDAP)

database, file-based exports of the case data will be created which will in turn be read by the Kea data integration scripts.

SECTION II - Attributes of the Data (use and accuracy)

1. Describe the uses of the data.

Data is used by SEC staff members to support investigative activities including conducting investigations of securities matters and coordinating cross-border (international) investigation and prosecution. Data also is used by SEC staff members to conduct examinations of regulated entities. Finally, data is used by SEC staff members to conduct research and analysis specifically in the area of analyzing market trading to identify and investigate alleged incidents of market abuse.

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: SEC Investigators (users) will use underlying data in existing investigations from the 9 data sources via EDAP as a flow-through tool for network analysis, clustering, enterprise data sharing/collaboration. Thus, the outcomes of the data analysis will naturally lead to new or broadened investigations of previously unknown patterns, concerns, etc. The new information collected about an individual or organization can be added to their existing investigation file/record or can lead to a new file being created. The information gleaned from data mining, clustering, network analysis, etc., could be used to take action against an individual(s) identified in an investigation using EDAP. If a new record is created, it would be made available to Government employees (including attorneys) who could make determinations against that individual (as EDAP has reporting capabilities with the data output easily shareable via file extracts, emails, presentations). Information would only be shared on a strict need to know basis and very selectively to authorized individuals. EDAP has a full audit trail so one can see who has access to data, when, and where it has been sent/shared.

3. How will the data collected from individuals or derived by the system be checked for accuracy?

Data would be checked for accuracy by SEC investigators or analysts if used in an investigation. SEC staff members may manually create, edit, or delete tags which relate disparate records, though an audit trail would remain. All of the information stored in EDAP is extra copies of records kept solely for convenience of reference.

SECTION III - Sharing Practices

1. Will the data be shared with any internal organizations?

No Yes If yes, please list organization(s): Enforcement (i.e Market Abuse, Forensic Accounting), Office of Compliance, Inspections, and Examinations (OCIE), based on existing staff access to data sources at the data element level.

2. Will the data be shared with any external organizations?

Privacy Impact Assessment

Palantir - Enterprise Data Analytics Platform (EDAP)

No Yes If yes, please list organizations(s): _____ How is the data transmitted or disclosed to external organization(s)? N/A. Data is not shared with external recipients.

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). The system accesses PII from various SEC systems that include: Tips, Complaints, and Referrals (TCR), Foreign Corrupt Practice Act (FSPA), ABAP (Bluesheet) Data, Accounting Fraud Data, and Filing Data (EDGAR). No SEC systems access ABAP.

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: EDAP does not collect data directly from individuals. Data is collected by other SEC source systems.

3. Do individuals have the right to consent to particular uses of the data?

Yes No N/A

Please explain: EDAP does not collect data directly from individuals. Data is collected by other SEC source systems.

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: Source data will continue to be stored in its native location and will not be loaded to the Palantir platform. Metadata created by the system and case specific datasets that are manually loaded to the system will be stored on local storage and maintained for seven years in accordance with current retention policies. Upon contract termination, data will be archived on removable media.

2. What are the procedures for identification and disposition of the data at the end of the retention period?

At the conclusion of the contract, the vendor is responsible for archiving (on CD) and returning all remaining SEC data to the Contracting Officer Representative and disposing of data in accordance with NIST SP 800-88.

Privacy Impact Assessment

Palantir - Enterprise Data Analytics Platform (EDAP)

3. Describe the privacy training provided to users, either generally or specifically relevant to the program or system?

All SEC staff and contractors receive annual privacy awareness training, which outlines their roles and responsibilities for properly handling and protecting PII.

4. Has a system security plan been completed for the information system(s) supporting the project?

Yes If yes, please provide date C&A was completed:

No If the project does not trigger the C&A requirement, state that along with an explanation.

The System Security Plan (SSP) was completed in November 2013 and was refreshed on December 19, 2013. In addition, the system is currently operational under a waiver.

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 (i.e. periodic, recurring, etc.) PII data extractions from the system?

No

Yes If yes, please explain: Nothing regular will be scheduled in Palantir to export PII information. Users may themselves export pieces of data, but this is entirely manually driven and is not regular or automated.

7. Which user group(s) will have access to the system?

The SEC TCR team, FCPA team, ABAP team, Accounting Fraud team, OCIE team, MicroCap team and the contractor team. Individuals do not have access to data in EDAP system unless they already have access to that data in the underlying system of record. Analysts and investigators (accountants and attorneys) would be prime users.

8. How is access to the data by a user determined? Palantir uses role-base and case-specific access control. The EDAP links with the SEC LDAP system and Active Directory to control access to data. User access control procedures have recently been completed detailing the process of accessing data, granting permission, and auditing.

Are procedures documented? Yes No

9. How are the actual assignments of roles and rules verified.

SEC Operating Directive - IT Security Human Resources Security Program is followed. Once a user leaves the organization they would have their access to EDAP terminated through AD and further, there are monthly reports of users generated by Palantir that will be reviewed for verification.

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

Palantir's Knowledge Management model further protects privacy and civil liberties by providing data transparency, immutable audit trails, and fine-grained security controls. Audit

Privacy Impact Assessment

Palantir - Enterprise Data Analytics Platform (EDAP)

trails include information about each time data is viewed, tagged, or exported including by whom and the time of the specific activity that occurred.

Palantir's fine-grained access control model is enforced throughout the entire Palantir application. It restricts what data users are able to view, down to individual data items. The permissions within Palantir are synced with the SEC's Active Directory permissions, and match the permission structure of the SEC's source system wherever possible. Without permission to access a given data item, a user will not be able to see that it exists, let alone access it. All search results are filtered on the server-side using the access control model, and no data will ever be transmitted to the application on the client's machine that they do not personally have permission to view.

SECTION VI - Privacy Analysis

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

Data is being protected through the implementation of the necessary security controls in accordance with the SEC Security Authorization Process.

During its early stages, user permissions were handled using a system internal to Palantir. This was easier to set up, but risked becoming out of date with the SEC's overall permission structure. The current version of EDAP integrates with the SEC's Active Directory for both authentication and data permission purposes. This ensures that only authorized users will receive access to any data within Palantir.

The EDAP system is now covered by detailed Audit and Accountability procedures concerning topics such as which records are audited at what frequency and by comparing to which external sources to ensure that actual data access conforms to authorized use of the system.