U.S. Department of Commerce
U.S. Patent and Trademark Office

Privacy Threshold Analysis
for the
Enterprise Data Services System – Databricks (EDS-DBX)

U.S. Department of Commerce Privacy Threshold Analysis

USPTO Enterprise Data Services System – Databricks (EDS-DBX)

Unique Project Identifier: EBPL-DA-03-00

Introduction: This Privacy Threshold Analysis (PTA) is a questionnaire to assist with determining if a Privacy Impact Assessment (PIA) is necessary for this IT system. This PTA is primarily based from the Office of Management and Budget (OMB) privacy guidance and the Department of Commerce (DOC) IT security/privacy policy. If questions arise or further guidance is needed in order to complete this PTA, please contact your Bureau Chief Privacy Officer (BCPO).

Description of the information system: *Provide a brief description of the information system.*
The E-Government Act of 2002 defines "information system" by reference to the definition section of Title 44 of the United States Code. The following is a summary of the definition: "Information system" means a discrete set of information resources organized for the collection, processing, maintenance, use, sharing, dissemination, or disposition of information. See: 44. U.S.C. § 3502(8).

The Information System will analyze datasets from the Big Data Reservoir-Trademark Quality Review (BDR-TQR) application to provide analytics and use the data for machine learning and Artificial Intelligence (AI) to improve business processes, the application achieves this using it's computing capabilities. Machine Learning (ML) is also used for fraud detection.

Databricks parses unique datasets from BDR-TQR to process data and deliver to a specified output location internal to USPTO, the system obtains more information out of the datasets it has been provided through analysis.

When there is an available amount of data to be processed, it is called and processed by the application then discarded after an output has been derived. The Databricks application works by using logic to parse information and derive an output.

The application provides a unified, open platform for USPTO contractors (data scientists, engineers and analysts) to write interactive and scheduled analysis workloads. USPTO employees (system owners and technical leads) also have access to the system to approve access requests before and administrator account is provisioned for contractors. Databricks is hosted on the Amazon Web Services (AWS) cloud platform, the data that is ingested into the application will be transferred from the BDR-TQR application. The OKTA (ICAM-IDaaS), QRadar and Environment-as-a-Service (EaaS) applications provide Identity Management, User Audit logging solutions and Email services respectively.

Address the following elements:

a) Whether it is a general support system, major application, or other type of system

Databricks is a Major Application hosted in the Amazon Web Services (AWS) cloud as a Software as a Service (SaaS) platform. The purpose of the interconnecting systems is to transfer data to be processed by Databricks, the transfer of data is done using a private link between each connecting system.

b) System location

The application is hosted in the AWS cloud.

c) Whether it is a standalone system or interconnects with other systems (identifying and describing any other systems to which it interconnects)

Databricks is interconnected with:

Big Data Reservoir (BDR)/Trademark Quality Review (TQR) – Both of these interconnections are covered under the Open Data-Big Master System (OD/BD/MS). TQR is a component of BDR. In addition to the BDR Portal, BDR also provides the TQR Portal. The TQR Portal is a single page application that harnesses modern web technologies and provides Trademark reviewers a single location to look at Trademark applications and the review processes. The TQR Portal provides quality reviewers with a centralized location to view the Dockets that are on the queue for review and additional features that include reviewing Trademark Review forms and completing necessary actions, final and non-final. The TQR Portal also includes reports that are generated using data that is captured by the BDR ingestion phase, it also provides supervisors the ability to view Trademark Reviews historical data interactions and list of Reviews completed within specified timeframes.

ICAM Identity as a Service (ICAM-IDaaS) – This application provides an enterprise authentication and authorization service to all applications/AIS's. As part of the enterprise services it will also provide compliance for some of the NIST 800-53 controls (e.g. AC, AU AP). The system provides following services to the enterprise: User Provisioning and Life Cycle Management, User Roles and Entitlement Management, User Authentication and Authorization to protected resources, Application Integration/Protection, NIST controls compliance related to SC and SI family.

QRadar – This application is a Security Information and Event Manager (SIEM) system that collects and consolidates USPTO Information System event log data from all USPTO hosts and devices configured to send their audit log files to the SIEM QRadar collector devices. The SIEM system collects network infrastructure, net-flow data to use in conjunction with operating system and hardware log data that it collects and stores. The SIEM system maintains separate data and correlates events to share with the EMS system to produce actionable, real-time alerts and around the clock monitoring in detailed display.

EaaS - The Enterprise Office Software Services (EOSS) product of the Enterprise Infrastructure Product Line (EIPL) provides communication and collaboration tools and services using Microsoft Office 365, OpenText facsimile and Blackberry COTS hosted in cloud and on-premise datacenters. Microsoft Office 365 component of EOSS provides support and is responsible for the core infrastructure of Office 365 including Exchange Online, SharePoint Online, OneDrive, and TEAMS. Each Service Team is responsible for the configuration and feature settings within their perspective Service.

d) The purpose that the system is designed to serve

The application provides a unified, open platform for data scientists, engineers and analysts to write interactive and scheduled analysis workloads.

e) *The way the system operates to achieve the purpose*

Databricks works by using logic to parse information from large datasets to derive an output. After the processing is completed the initial data is discarded and the output is sent to a specified location, Databricks is called every time a batch of data is ready to be processed.

f) A general description of the type of information collected, maintained, used, or disseminated by the system

Several data elements are analyzed which may include names, addresses and patent data, the system will not process, store or transmit any sensitive PII.

g) Identify individuals who have access to information on the system

Only the administrators have access to data in the application, Databricks is not a public facing website. The data in the application can be retrieved using scheduled jobs and SQL queries.

h) How information in the system is retrieved by the user

Transfer of data into Databricks for processing is done using a private link between each connecting system.

i) How information is transmitted to and from the system

The information is transmitted using TLS 1.2 via bulk data transfer from other systems internal to USPTO.

Questionnaire:

1. Status of the Information System
1a. What is the status of this information system?

☒ This is a new information system. *Continue to answer questions and complete certification.*

☐ This is an existing information system with changes that create new privacy risks. *Complete chart below, continue to answer questions, and complete certification.*

Changes That Create New Privacy Risks (CTCNPR)					
a. Conversions	☐	d. Significant Merging	☐	g. New Interagency Uses	☐
b. Anonymous to Non-Anonymous	☐	e. New Public Access	☐	h. Internal Flow or Collection	☐
c. Significant System Management Changes	☐	f. Commercial Sources	☐	i. Alteration in Character of Data	☐
j. Other changes that create new privacy risks (specify):					

☐ This is an existing information system in which changes do not create new privacy risks, and there is not a SAOP approved Privacy Impact Assessment. *Continue to answer questions and complete certification.*

☐ This is an existing information system in which changes do not create new privacy risks, and there is a SAOP approved Privacy Impact Assessment. *Skip questions and complete certification.*

1b. Has an IT Compliance in Acquisitions Checklist been completed with the appropriate signatures?

☒ Yes. This is a new information system.

☐ Yes. This is an existing information system for which an amended contract is needed.

☐ No. The IT Compliance in Acquisitions Checklist is not required for the acquisition of equipment for specialized Research and Development or scientific purposes that are not a National Security System.

☐ No. This is not a new information system.

2. Is the IT system or its information used to support any activity which may raise privacy concerns?

NIST Special Publication 800-53 Revision 4, Appendix J, states "Organizations may also engage in activities that do not involve the collection and use of PII, but may nevertheless raise privacy concerns and associated risk. The privacy controls are equally applicable to those activities and can be used to analyze the privacy risk and mitigate such risk when necessary." Examples include, but are not limited to, audio recordings, video surveillance, building entry readers, and electronic purchase transactions.

☐ Yes. *(Check all that apply.)*

Activities			
Audio recordings	☐	Building entry readers	☐
Video surveillance	☐	Electronic purchase transactions	☐
Other (specify):			

☒ No.

3. Does the IT system collect, maintain, or disseminate business identifiable information (BII)?

As per DOC Privacy Policy: "For the purpose of this policy, business identifiable information consists of (a) information that is defined in the Freedom of Information Act (FOIA) as "trade secrets and commercial or financial information obtained from a person [that is] privileged or confidential." (5 U.S.C.552(b)(4)). This information is exempt from automatic release under the (b)(4) FOIA exemption. "Commercial" is not confined to records that reveal basic commercial operations" but includes any records [or information] in which the submitter has a commercial interest" and can include information submitted by a nonprofit entity, or (b) commercial or other information that, although it may not be exempt from release under FOIA, is exempt from disclosure by law (e.g., 13 U.S.C.)."

☒ Yes, the IT system collects, maintains, or disseminates BII.

☐ No, this IT system does not collect any BII.

4. Personally Identifiable Information (PII)

4a. Does the IT system collect, maintain, or disseminate PII?

As per OMB 17-12: "The term PII refers to information that can be used to distinguish or trace an individual's identity either alone or when combined with other information that is linked or linkable to a specific individual."

☒ Yes, the IT system collects, maintains, or disseminates PII about: *(Check all that apply.)*

☒ DOC employees
☒ Contractors working on behalf of DOC
☐ Other Federal Government personnel
☒ Members of the public

☐ No, this IT system does not collect any PII.

If the answer is "yes" to question 4a, please respond to the following questions.

4b. Does the IT system collect, maintain, or disseminate Social Security numbers (SSNs), including truncated form?

☐ Yes, the IT system collects, maintains, or disseminates SSNs, including truncated form.

Provide an explanation for the business need requiring the collection of SSNs, including truncated form.
Provide the legal authority which permits the collection of SSNs, including truncated form.

☒ No, the IT system does not collect, maintain, or disseminate SSNs, including truncated form.

4c. Does the IT system collect, maintain, or disseminate PII other than user ID?

☒ Yes, the IT system collects, maintains, or disseminates PII other than user ID.

☐ No, the user ID is the only PII collected, maintained, or disseminated by the IT system.

4d. Will the purpose for which the PII is collected, stored, used, processed, disclosed, or disseminated (context of use) cause the assignment of a higher PII confidentiality impact level?

Examples of context of use include, but are not limited to, law enforcement investigations, administration of benefits, contagious disease treatments, etc.

☐ Yes, the context of use will cause the assignment of a higher PII confidentiality impact level.

☒ No, the context of use will not cause the assignment of a higher PII confidentiality impact level.

If any of the answers to questions 2, 3, 4b, 4c, and/or 4d are "Yes," a Privacy Impact Assessment (PIA) must be completed for the IT system. This PTA and the SAOP approved PIA must be a part of the IT system's Assessment and Authorization Package.

CERTIFICATION

☒ The criteria implied by one or more of the questions above **apply** to the Enterprise Data Services System – Databricks (EDS-DBX) and as a consequence of this applicability, a PIA will be performed and documented for this IT system.

☐ The criteria implied by the questions above **do not apply** to the Enterprise Data Services System – Databricks (EDS-DBX) and as a consequence of this non-applicability, a PIA for this IT system is not necessary.

System Owner Name: Scott Beliveau Office: Enterprise Advanced Analytics Branch (I/EAAB) Phone: (571) 272-7343 Email: Scott.Beliveau@uspto.gov Signature: ________________________________ Date signed: ______________________________	**Chief Information Security Officer** Name: Don Watson Office: Office of the Chief Information Officer (OCIO) Phone: (571) 272-8130 Email: Don.Watson@uspto.gov Signature: ________________________________ Date signed: ______________________________
Privacy Act Officer Name: Ezequiel Berdichevsky Office: Office of General Law (O/GL) Phone: (571) 270-1557 Email: Ezequiel.Berdichevsky@uspto.gov Signature: ________________________________ Date signed: ______________________________	**Bureau Chief Privacy Officer and Authorizing Official** Name: Henry J. Holcombe Office: Office of the Chief Information Officer (OCIO) Phone: (571) 272-9400 Email: Jamie.Holcombe@uspto.gov Signature: ________________________________ Date signed: ______________________________
Co-Authorizing Official Name: N/A Office: N/A Phone: N/A Email: N/A Signature: ________________________________ Date signed: ______________________________	