

U.S. Department of Commerce
U.S. Census Bureau



Privacy Threshold Analysis
for the
Enterprise Data Lake

U.S. Department of Commerce Privacy Threshold Analysis

U.S. Census Bureau/ Enterprise Data Lake

Unique Project Identifier: FISMA (CSAM) ID 2735

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 and its purpose: *Provide a general description of the information system in a way that a non-technical person can understand.*

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).

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

The Office of the Chief Information Officer (OCIO) Enterprise Data Lake (EDL) is a major application that will provide a Platform-as-a-Service (PaaS) for IT system and data owners, hosting all U.S. Census Bureau surveys data, from collection to tabulation, administrative records, and third-party data. The consolidation of survey processing systems will help the U.S. Census Bureau (USCB) to sustain its place as a leader in statistical methodologies and data products.

(b) System location

The Enterprise Data Lake will reside on the Amazon Web Services (AWS) GovCloud environment. The GovCloud environment is dispersed across two regions: US- East and US-West. AWS headquarters is located in Seattle, Washington. AWS and the Enterprise Data Lake perform data backups, that are stored at Census Bureau’s datacenter located in Bowie, MD.

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

EDL utilizes and interconnects with Cloud Services account for Amazon Web Services (AWS). EDL also interconnects with the IT systems that provide the common security policies (i.e., access and authorization controls) including Data Communications, Network Services, Enterprise Applications, and Office of Information Security (OIS) Systems. Connections will be established with the U.S. Census Bureau data collection,

processing, and dissemination systems in the survey program areas as needed for the ingestion, on-boarding, and dissemination preparation of data into EDL. EDL will receive survey data, via a temporary connection, from many of the U.S. Census Bureau data collection and processing systems including Economic Census and Surveys and Special Processing, Geography, Decennial, Demographic Census, Surveys and Special Processing, Center for Enterprise Dissemination, American Community Survey Office (ACSO), Foreign Trade Division Applications, EAD Windows Applications System, etc.

(d) The purpose that the system is designed to serve

The EDL is part of the Census Bureau's transformation efforts to become a data-centric organization that produces data faster, invests in new data products and collection methods. EDL serves as the enterprise data storage and computing platform that includes survey operation support, concurrent and research analytics, post processing, product creation, product innovation, and archiving.

(e) The way the system operates to achieve the purpose(s) identified in Section 4

The IT system was created to support the Census Bureau's longstanding leadership in data analytics and technology. The EDL makes use of AWS Infrastructure-as-a-Service (IaaS), which is a form of cloud computing that provides computing resources over the internet. EDL provides a PaaS and Software as a Service (SaaS) to its Census Bureau program areas seeking to consolidate data analytics, data management and data storage activities. EDL consolidation of IT systems will allow program areas to leverage standardized data services to centrally govern the programs' data to deliver timely, consistent, and accurate data products.

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

Survey data that is ingested and maintained by EDL is provided by the Census Bureau program areas and the survey owners in respective survey areas that are conducting mission-related studies. These surveys include PII and BII from members of the United States public, which may include federal employees and contractors, and personnel of business entities. The data will be used by EDL users to provide survey data management, processing, analytics, and storage services to Census Bureau program areas.

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

U.S. Census Bureau employees, contractors, and researchers under Sworn Status can access EDL environment. Only authorized users with a need-to-know basis can retrieve data in EDL.

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

Census Bureau Program data stakeholders will use an interface based on appropriate access and control rights. The user will be able to retrieve the information using a web-based user interface to access or process the data based on his/her role and permissions. Via the web-based user interface, the user can spin up computing environments and load data. The interface will trigger several data processing activities such as batch and interactive processing. This capability provides authorized users visibility into the data transformations that occur, as raw uploaded data moves to the final data product that is published and used for research.

Users with a need to know are able to retrieve data in EDL by unique identifiers.

(i) How information is transmitted to and from the system

EDL receives survey data from the IT systems that are utilized by Census Bureau survey program areas. The program area IT systems will upload survey data from present and past surveys and administrative records data where the Social Security Number (SSN) has been replaced with a unique non-identifying code called a protected identification key (PIK). The data ingested is stored in the storage layer of the EDL and made available for program area consumption after the necessary access approvals are obtained. The EDL will make use of secure tools to ingest data.

(j) Any information sharing conducted by the system

The EDL system makes data available to Census Bureau program areas that have been authorized to conduct data analytics and management activities involving the information collected from internal census survey systems. Data Communications, Network Services, Enterprise Applications, and Office of Information Security (OIS) Systems.

Questionnaire:

1. What is the status of this information system?

 X 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 (version 01-2015 or later). *Skip questions and complete certification.*

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.

 X Yes. *Please describe the activities which may raise privacy concerns.*

Audio Recordings

 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.)."

 X 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

☐ National Institute of Standards and Technology Associates

☒ 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.

EDL is a repository system, SSN could reside within survey data in the EDL. Other PIAs for data hosted on EDL will contain SSN justifications, as applicable. For example, depending on the scope of the survey, Demographic surveys may need to collect SSN's from respondents. If collected, SSN's are treated as PII and subjected to the protections under Title 13.

☐ 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.

- 4c. 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, and/or 4c are “Yes,” a Privacy Impact Assessment (PIA) must be completed for the IT system. This PTA and the approved PIA must be a part of the IT system’s Assessment and Authorization Package.

CERTIFICATION

 X I certify the criteria implied by one or more of the questions above **apply** to the Enterprise Data Lake and as a consequence of this applicability, I will perform and document a PIA for this IT system.

 I certify the criteria implied by the questions above **do not apply** to the [IT SYSTEM NAME] and as a consequence of this non-applicability, a PIA for this IT system is not necessary.

Name of Information System Security Officer (ISSO): Kolawole O. Rotimi Jr.

Oluwagbemiga
Signature of ISSO: Seun Adekunle

Digitally signed by Oluwagbemiga Seun Adekunle
 Date: 2021.12.07 14:50:21 -05'00' Date: _____

Name of System Owner (SO): Gregg “Skip” Bailey

KENNETH
Signature of SO: HARRISON

Digitally signed by KENNETH HARRISON
 Date: 2021.12.07 14:25:55 -05'00' Date: _____

Name of Chief Information Security Officer (CISO): Beau Houser

BEAU HOUSER
Signature of CISO:

Digitally signed by BEAU HOUSER
 Date: 2021.12.07 12:52:01 -05'00' Date: _____

Name of Authorizing Official (AO): Luis J Cano

LUIS CANO
Signature of AO:

Digitally signed by LUIS CANO
 Date: 2021.12.06 13:38:39 -05'00' Date: _____

Name of Bureau Privacy Officer (BPO): Byron Crenshaw

BYRON
Signature of BPO: CRENSHAW

Digitally signed by BYRON CRENSHAW
 Date: 2021.12.07 11:50:59 -05'00' Date: _____

Name of Privacy Act Officer (PAO): Byron Crenshaw

BYRON
Signature of PAO: CRENSHAW

Digitally signed by BYRON CRENSHAW
 Date: 2021.12.07 11:51:21 -05'00' Date: _____