Template Version Number: 01-2021

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

Privacy Threshold Analysis
for the
Office of the Chief Information Office (OCIO)
Chief Technology Office (CTO)
Enterprise Data Lake (EDL)

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: *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).

Office of the Chief Information Officer (OCIO) Chief Technology Office (CTO) 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.

EDL is an enterprise-wide, big data management platform that modernizes data storage and data analysis capabilities across all its directorates with appropriate role-based access control. EDL supports the Census Bureau's data and analytical needs in a secure, scalable, high-performing storage and computing cloud environment with appropriate backups to the Census datacenter. This platform increases the Census Bureau's capability to ingest the ever-increasing volume of administrative records, improve the quality of data products and apply disclosure avoidance to protect PII data as required by Title 13, Title 26, and other data protection laws. The EDL will fully support the Process, Derive, and Publish stages of the data lifecycle.

Address the following elements:

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

Enterprise Data Lake is a major application.

b) System location

The Enterprise Data Lake resides on the Amazon Web Services (AWS) GovCloud environment. The GovCloud environment is dispersed across two regions: US- East and US-West. AWS headquarters is in Seattle, Washington. AWS and the EDL 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 enterprise security systems for access and authorization controls such as OCIO Data Communications, OCIO Network Services, OCIO Enterprise Applications, and OCIO Office of Information Security (OIS) Systems. EDL will host data for the Decennial Census, Economic Programs, Demographic Surveys, and the American Community Survey.

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

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, used, 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, BII, and FTI 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 can retrieve data in EDL by unique identifiers (i.e. JBID, username and password).

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. EDL cloud service providers do not have access to the encryption keys of Census Bureau data therefore they do not have access to the data.

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.*

_X__ 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): Addition of FTI data					

____ 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.

__X__ 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.

_X_ Yes. *(Check all that apply.)*

Activities			
Audio recordings	X	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.)."

_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."

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

_X_ DOC employees
_X_ Contractors working on behalf of DOC
_X_ Other Federal Government personnel
_X_ 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.

_X___ 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?

_X_ 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.

_X_ 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

__X__ The criteria implied by one or more of the questions above **apply** to the OCIO CTO EDL 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 [IT SYSTEM NAME] and as a consequence of this non-applicability, a PIA for this IT system is not necessary.

System Owner Name: Gregg Bailey Office: Office of the Chief Information Officer Phone: 301-763-0989 Email: gregg.d.bailey@census.gov Signature: Gregg Bailey Digitally signed by Gregg Bailey Date: 2023.02.14 09:26:54 -05'00' Date signed:	**Chief Information Security Officer** Name: Beau Houser Office: Office of Information Security Phone: 301-763-1235 Email: beau.houser@census.gov Signature: BEAU HOUSER Digitally signed by BEAU HOUSER Date: 2023.02.21 08:21:39 -05'00' Date signed:
Privacy Act Officer Name: Byron Crenshaw Office: Policy Coordination Office Phone: 301-763-7997 Email: byron.crenshaw@census.gov Signature: BYRON CRENSHAW Digitally signed by BYRON CRENSHAW Date: 2023.03.01 14:08:13 -05'00' Date signed:	**Authorizing Official** Name: Luis J. Cano Office: Office of the Chief Information Officer Phone: 301-763-3968 Email: luis.j.cano@census.gov Signature: LUIS CANO Digitally signed by LUIS CANO Date: 2023.02.23 10:18:21 -05'00' Date signed:
Bureau Chief Privacy Officer Name: Byron Crenshaw Office: Policy Coordination Office Phone: 301-763-7997 Email: byron.crenshaw@census.gov Signature: BYRON CRENSHAW Digitally signed by BYRON CRENSHAW Date: 2023.03.01 14:08:34 -05'00' Date signed:	**Business Authorizing Official** Name: Nick Orsini Office: Associate Directorate for Economic Programs Phone: 301-763-6959 Email: nick.orsini@census.gov Signature: NICK ORSINI Digitally signed by NICK ORSINI DN: c=US, o=U.S. Government, ou=Department of Commerce, ou=Bureau of the Census, cn=NICK ORSINI, 0.9.2342.19200300.100.1.1=13001000014152 Date: 2023.02.28 10:24:17 -05'00' Date signed: