PRIVACY IMPACT ASSESSMENT

OBO Azure Data Lake PIA

1. Contact Information

A/GIS Deputy Assistant Secretary Bureau of Administration Global Information Services

2. System Information

(a) Date of completion of this PIA: February 2023
(b) Name of system: OBO Azure Data Lake
(c) System acronym: OADL
(d) Bureau: Overseas Buildings Operations (OBO)
(e) iMatrix Asset ID Number: 319010
(f) Child systems (if applicable) and iMatrix Asset ID Number: N/A

(g) Reason for performing PIA:

- ☒ New system
- ☐ Significant modification to an existing system
- ☐ To update existing PIA for a triennial security reauthorization

(h) Explanation of modification (if applicable): Click or tap here to enter text.

3. General Information

(a) Does the system have a completed and submitted data types document in Xacta?

☒Yes
☐No - Contact IRM/IA at IASolutionCenter@state.gov for assistance.

(b) Is this system undergoing an Assessment and Authorization (A&A)?

☒Yes
☐No

If yes, has the privacy questionnaire in Xacta been completed?
☒Yes
☐No

(c) Describe the purpose of the system:

The OBO Azure Data Lake (OADL) system provides an OBO-governed hybrid cloud data ecosystem which creates a centralized integration data repository for OBO systems, such that all current and future reporting, dashboarding and analytics requirements are

catered in this system. Information is collected from OBO Real Property Application (OBO RPA), Global Maintenance Management System (GMMS), Building Management Information System (BMIS) and Safety Health and Environmental Management (SHEM) Risk Management System (SRMS).

The system will support the business integration requirement of data and ensure timely availability of data for all of OBO's analytical needs. OADL will also allow information to be shared with OBO senior management via dashboards and reports generated by OBO business analysts and Data Scientists, so that OBO is able to make data driven decisions.

OADL has 2 containers:

1. Master Data Lake (MDL)
   Data in the MDL is a mirror copy of the data from the corresponding source system it was captured and may include PII.
2. Validated Data Lake (VDL)
   The end users/consumers seeking access to OADL are only granted access to the VDL. There is no unmasked PII data in the VDL unless the end user/consumer have a specific purpose that requires the use of PII.

(d) Describe the personally identifiable information (PII) that the system collects, uses, maintains, or disseminates:

OADL contains PII collected by other OBO systems:

BMIS

- Names of Primary Occupants
- Birthdates of Primary Occupants
- Phone number(s) of Primary Occupants
- Personal Address of Post Assigned Resident
- Financial Information
- Personnel Information

RPA

- Names of Primary Occupant
- Birthdates of Secondary Occupant
- Personal Address of Post assigned residence
- Phone number(s) of Primary Occupant
- Property Owner Name – (US persons/non-US persons)
- Property Owner Personal Email Address
- Property Owner Personal Address
- Property Owner Personal Phone Number
- Financial Information
- Personnel Information

SRMS

- Name (of the person involved with the mishap);
- Date of birth (of the person involved with the mishap if an employee)
- Medical Information (Nature of injury or occupational illness such as lacerations, heat stress, etc. and affected body parts (e.g., knees, leg, arm))

GMMS does not collect PII.

(e) What are the specific legal authorities and/or agreements that allow the information to be collected?

22 U.S.C. § 300 - Dispositions of property; damage payments; acceptance of gifts or services
5 U.S.C. 5912 - Quarters in Government owned or rented.

(f) Is the information searchable by a personal identifier (e.g., name or Social Security number, etc.)?

☒Yes, provide:

- SORN Name and Number:
  Contractors Records, State-45
  Human Resources Records, State-31

- SORN publication date (found under the Volume Number and above the Public Notice Number on the published SORN):
  September 27, 1977
  July 19, 2013

☐No, explain how the information is retrieved without a personal identifier.

(g) Does the existing SORN need to be amended to reflect the inclusion of this new or significantly modified system? ☐Yes ☒No

If yes, please notify the Privacy Office at Privacy@state.gov.

(h) Is there a records retention schedule submitted to or approved by the National Archives and Records Administration (NARA) for this system? ☒Yes ☐No
(If uncertain about this question, please contact the Department's Records Officer at records@state.gov.)

If yes provide (Consolidate as much as possible):

- Schedule number (e.g., (XX-587-XX-XXX)):
  N/A

- Disposition Authority Number:
  DAA-0059-2020-0019-0011

- Length of time the information is retained in the system:
  Destroy 7 years after the property is no longer owned/leased by the Department.

- Type of information retained in the system:
  Contains data regarding costs, leases, and other contracts associated with real estate and facilities; project management tracking; planning and budget data; Property and leasing records, post/consulate projects that have financial impact, such as repairs & construction, scope, schedule, and budget. OADL also collects and maintains mishaps investigation reports.

4. Characterization of the Information

(a) What entities below are the original sources of the information in the system? Please check all that apply.

☒ Members of the Public
☒ U.S. Government employees/Contractor employees
☒ Other (people who are not U.S. Citizens or LPRs)

(b) On what other entities above is PII maintained in the system?

☐ Members of the Public
☐ U.S. Government employees/Contractor employees
☐ Other
☒ N/A

(c) If the system contains Social Security Numbers (SSNs), is the collection necessary?

☐Yes ☒No ☐N/A

- If yes, under what authorization?

(d) How is the PII collected?

Information in OADL is extracted from other OBO source systems. The Informatica Data Quality (IDQ) tool extracts the source information from the respective source systems and loads it into OADL. The PII is included in the source information captured, along with all other data being extracted.

Note: Our process will run as a downstream to all/any OBO systems and will connect with the corresponding databases to extract the data from them.

(e) Where is the information housed?

☐ Department-owned equipment
☒ FEDRAMP-certified cloud

☐ Other Federal agency equipment or cloud
☐ Other

- If you did not select "Department-owned equipment," please specify.
  The information is housed in the Department's FEDRAMP approved Enterprise Azure Cloud environment.

(f) What process is used to determine if the PII is accurate?

The Informatica Data Quality (IDQ) mappings run as a downstream daily batch process and extracts the source information in its as-is form. Any checks for PII accuracy occur on the respective source systems, and OADL relies on source systems to provide accurate information. The accuracy of PII is covered in the systems respective PIAs.

(g) Is the information current? If so, what steps or procedures are taken to ensure it remains current?

Yes. OADL relies on source systems to provide current information. The IDQ mapping is also scheduled to run daily batch to ensure the information in OADL is current.

(h) Does the system use information from commercial sources? Is the information publicly available?

No, the system does not use information from commercial sources, and it does not use any information that is publicly available.

(i) How was the minimization of PII in the system considered?

OADL contains two repositories which include the Master Data Lake (MDL) and the Validated Data Lake (VDL). All data captured in OADL must be approved by OBO management prior to being pulled from source systems into the MDL, which is a highly restricted environment accessible to only system and database administrators. This raw data sourced from source systems may include PII on occasions but will not be collected if not required to support the business needs. End users including OBO approved data scientists, dashboard developers, and other data analyst are never granted access to the Master Data Lake, but rather the VDL which contains a subset of the raw data sourced from the source systems. PII transferred to the Validated Data Lake and accessible to aforementioned end users is limited by Role Based Access Controls (RBAC) or in some instances masked to further regulate access.

5. Use of information

(a) What is/are the intended use(s) for the PII?

The intended use of the information in OADL is to meet and cater to all current and future OBO reporting and dashboarding requirements. The PII is included in the source information captured, along with all other data being extracted.

(b) Is the use of the PII relevant to the purpose for which the system was designed or for which it is being designed?

Yes. The purpose of the OBO Azure Data Lake (OADL) is to enable OBO to be able to build analytics, reports, and dashboards. In addition, OADL will have data from all source systems including PII in the OADL so that the data scientist/analysts from each individual source system are able to better analyze and identify other Key Performance Indicators (KPI's) for reports/dashboards etc.

(c) Does the system analyze the PII stored in it? ☐Yes ☒No

If yes:

(1) What types of methods are used to analyze the PII?

(2) Does the analysis result in new information?

(3) Will the new information be placed in the individual's record? ☐Yes ☐No

(4) With the new information, will the Department be able to make new determinations about the individual that would not have been possible without it? ☐Yes ☐No

(d) If the system will use test data, will it include real PII?

☒Yes ☐No ☐N/A

If yes, please provide additional details.

OBO currently tests processes on Enterprise Azure Cloud Service (EACS) stage which is built in EACS production. Data is not moving out of the EACS production environment. Both Test and production environments for the OBO Azure Data Lake will collect data from production source systems only.

6. Sharing of PII

(a) With whom will the PII be shared internally and/or externally? Please identify the recipients of the information.

Internal:
OADL does not share PII with other Department bureaus/offices.

External:
OADL does not share PII outside of the Department.

(b) What information will be shared?

Internal:
N/A

External:
N/A

(c) What is the purpose for sharing the information?

Internal:
N/A

External:
N/A

(d) The information to be shared is transmitted or disclosed by what methods?

Internal:
N/A

External:
N/A

(e) What safeguards are in place for each internal or external sharing arrangement?

Internal:
N/A

External:
N/A

7. Redress and Notification

(a) Is notice provided to the record subject prior to the collection of his or her information?

No, OADL does not collect any PII directly from any record subjects, all data is extracted from source systems. Where applicable, those systems will provide notice.

(b) Do record subjects have the opportunity to decline to provide the PII or to consent to particular uses of the PII?

☐Yes ☒No

If yes, how do record subjects grant consent?

If no, why are record subjects not allowed to provide consent?

OADL does not collect the information directly from the record subject, therefore there is no opportunity to allow the provision of consent.

(c) What procedures allow record subjects to gain access to their information?

OADL does not provide access for record subjects. Access to information may be possible within source systems.

(d) Are procedures in place to allow a record subject to correct inaccurate or erroneous information?

☐Yes ☒No

If yes, explain the procedures.

If no, explain why not.

OADL does not allow record subjects to gain access to their information within OADL.

(e) By what means are record subjects notified of the procedures to correct their information?

OADL does not allow record subjects to correct their information. Records subjects may be able to correct their information within source systems by following procedures laid out by those source systems.

Record subjects can also follow procedures laid out in the applicable SORNs noted in 3f.

8. Security Controls

(a) How is all of the information in the system secured?

The operation system, database, and Azure Government cloud are configured to DS security configuration standards written guidelines on how to set up a system appropriately and securely. In addition, the application uses PIV card authentication/Single Sign-On (SSO). Security tools such as Splunk, Rapid 7, Oracle Database Appliance (ODA), and Symantec Endpoint Protection, are in place to ensure compliance, vulnerability management, and a proper continuous monitoring strategy to ensure the confidentiality, integrity, and availability of the system and its information.

(b) Explain the different roles that have been created to provide access to the system and the PII (e.g., users, managers, developers, contractors, other).

Administrators - Administer the OADL environment and ensure access to the data in the data lake is given only to approved users. The consumers seeking access to the data lake will have to submit a ticket in the OBO system, stating the purpose for it clearly and the ticket will be routed to all the appropriate stake holders for approvals. Administrators will have access to all PII that is present in OADL including the Master Data Lake.

Read Only Users - These are READ only roles. This role does not have access to unmasked PII data currently, however if they had a specific business need to use PII, they would be able to request access to specific data sets that may include any PII in 3d. OADL currently has 11 folders in the Validated Data Lake and each of these are governed by a corresponding role-based access control, meaning that a Read Only User will only have access to specific folders of data within the system, and will not automatically have access to all data in the Validated Data Lake.

(c) Describe the procedures established to limit system and data access to only those individuals who have an "official" need to access the information in their work capacity.

For access to the VDL, management's written approval is required, and approval is based on position as well as a need-to know. Audits are completed annually via User Validation by post and OBO Headquarters to remove access to the system of personnel who no longer have a need-to-know. The end users/consumers seeking access to the data lake are only granted access to the VDL. In order to obtain access to the VDL and specific role-based containers within the VDL, the user will have to submit a written request in OBO system and must be approved by an OBO authorized approver.

The access to MDL is restricted only to System Administrator. System Administrator roles are provisioned by IRM/OPS/SIO per their employment and business needs. IRM/SIO are the Azure Government tenant (USDOSMAG) owners in which the Azure Data Lake resides. OBO has a subscription within the SIO Azure Government tenant (USDOSMAG)

OBO security groups have been granted a certain level of access which was previously agreed upon by the OBO ISSO and the SIO ISSO. The access that has been granted to Azure resourced for the OBO administrators is always of least privileges per Microsoft Azure best practices.

All access is granted and tracked by the Enterprise or rOBO ServiceNow ticket system with an approved request. First step in the process for access being granted to the OBO administrator(s) is the submittal of an rOBO account request and/or an additional access request form ticket. The ticket then goes through approval the process, first by the administrator's government manager (GTM) and final approval lays with the OBO ISSO.

After all approvals, the OBO administrators will take for action and add new administrators to existing security groups mentioned above which grant access to specific elements of Azure (i.e., ADLS).

(d) How is access to data in the system determined for each role identified above?

The OBO Office of Planning Real Estate (PRE) determines access based on need and approved level of access for each role listed in 8b. Access request to folders is based on least privilege as supported by the business need. For example, an individual working on an RPA related report/analysis will be granted access to only RPA folder by adding them to MAG OBO ADLS RPA.

In order to gain access as an Administrator, an employee must be part of the Network Operations Management Branch (NOMB) group, who are usually responsible for administrative related tasks within any OBO environment including EACS, AZURE, ADLS, etc.

In order to gain access as a Read Only User, an employee must be approved by the respective Source System Owner as well. They must also have an approved OBO ticket with a clearly stated purpose behind seeking access to the corresponding folder.

(e) What monitoring, recording, auditing safeguards, and other controls are in place to prevent the misuse of the information?

NIST 800-53 Rev 4, App. J controls are in effect and enforced, and auditable events are captured by Splunk. The information can only be accessed by PIV card and is Single Sign-on enabled which prevents impersonation by any users. Changes or additions to existing data are captured in application audit logs. The audit logs document unintended modification and unauthorized access, and dynamically audit retrieval access to data that are designated as critical.

(f) Are procedures, controls, or responsibilities regarding access to data in the system documented?

☒Yes ☐No

(g) Explain the privacy training provided to each role identified in 8(b) that has access to PII other than their own.

All OpenNet users with access to OADL are required to take the mandatory PS800 - Cybersecurity Awareness Training, which has a privacy component, annually and the PA318 “Protecting Personally Identifiable Information” training on a biennial basis. There are no additional role-based privacy trainings.