Privacy Impact Assessment

Enterprise Data Analytics Platform & Toolset (EDAPT)

- Version: 3.0
- Date: April 30, 2021
- Prepared for: USDA Office of Chief Information Officer (OCIO)

United States Department of Agriculture

Privacy Impact Assessment for the

Enterprise Data Analytics Platform & Toolset (EDAPT)

April 30, 2021

Contact Point

Ted Kaouk
Office of the Chief Information Officer
202-690-7306

Reviewing Official

Ted Kaouk
United States Department of Agriculture
(202) 690-7306

Abstract

This Privacy Impact Analysis (PIA) is for Enterprise Data Analytics Platform & Toolset (EDAPT) of the Office of Chief Information Officer (OCIO). EDAPT is designed to collect, stores data from USDA agencies systems to provide aggregated data in multiple dashboard and visualizations and provide a subset of users the ability to perform self-service analytics using the data. This PIA is being updated based on annual A&A and to documents privacy protections incorporated in EDAPT.

Overview

The USDA EDAPT is an enterprise analytics environment that provides a holistic view of business activities, all of which aim to take a data-driven approach to the services USDA provides customers.

The system provides the following capabilities:

- Executive Dashboards – Executive level dashboards across administrative functions: IT, HR, Finance, Property and Fleet, Contracting and Procurement, Operations, and Homeland Security.
- Mission Area & Departmental Dashboards – Mission area and department-specific dashboards to provide key metrics across programs and initiatives as well as for Public use. Enterprise Analytics - Administrative, Mission Area, and Public Data – Stores and makes available data from across the administrative areas, mission areas, and public data sources for answering cross cutting analytic questions and deliver sophisticated, data driven insights.
- Custom Applications
    a. SRA custom application collects non-PII budget data that is input by users of the application for transactional purposes.
    b. Payment Schedule custom application collects scenario and cost component data that is input by users of the application for transactional purposes.
    c. The DRUID application is currently in development and will provide a Web UI and API. The web app will provide analytics (reports, stats, dashboards) and extract access via API. This is not a “data-entry” application.
- Citrix enabled virtual environment used for DSW and NASS IMAGES to provide users with high-performance compute resources and advanced analytical tools to enable data scientists to perform statistical analysis and business intelligence.
- Informatica EDC/Axon for analysis and governance of metadata.

Section 1.0 Characterization of the Information

The following questions are intended to define the scope of the information requested and/or collected as well as reasons for its collection as part of the program, system, rule, or technology being developed.

1.1 What information is collected, used, disseminated, or maintained in the system?

EDAPT will house data from eight USDA Mission Areas and six USDA offices (including data regarding USDA employees, contractors, borrowers, grantees, recipients and publicly available data sources). EDAPT will house many categories of data from numerous data sources. This may include but is not limited to Personally Identifiable Information (PII) personal data, work–related data, geolocation data and data on USDA employees, grantors, contractors.

1.2 What are the sources of the information in the system?

Information contained in this system is obtained from USDA and publicly available federal source systems:

1. 1OCIO
2. ACAAZ (and child systems)
    a. ABI
    b. BIIS
    c. CEMS
    d. DBIGS
    e. ePVP
    f. eTDE
    g. FEIRS
    h. MNIS
    i. MNP
    j. OID
    k. SESIS
    l. SLIMS
3. ACE
4. AgLearn (has PII)
5. AgMax
6. AIP (Interconnection agreement/ No Data)
7. ARM (has PII)
8. ARSnet
9. CONCUR (has PII)
10. CORE SERVICES (has PII)
11. CPAIS (has PII)
12. CSAM

13. eAuth
14. eDRS (has PII)
15. E-GSS (FSIS) (has PII)
16. EPACS (has PII)
17. eWCMD
18. EWPPT
19. FMMI (has PII)
20. FPAC Salesforce (has PII)
21. FPAC ServiceNow (has PII)
22. FSIMS (has PII)
23. GMRS
24. IAS
25. iMart (has PII)
26. MDA-L
    a. ATT
    b. FAIS
    c. GAIN
    d. GATS
    e. PSD
    f. SUGARS
    g. STAR
    h. UES
27. MFIS – PAS RD (has PII)
28. MIDAS
29. Midrange Systems
30. MRP AWS GSS - VS DIS
31. NASS Data Applications
32. NFC Insight (has PII)
33. OSCAR (has PII)
34. P-GSS (FNS) (has PII)
35. ProTracts-FundManager
36. RD EDW (has PII)
    a. RD TDW
37. ROD-CS
38. SES
39. SNOW ISC
40. Telecom
41. USA Staffing (has PII)
42. VSISM (has PII)
43. WBSCM
44. WIMS
45. AMS MyMarketNews – Mars Cotton (Public Data)
46. Census (Public Data)
47. Discovery Labs (Public Data)
48. DOE Eagle-I - DOE Power Outage (Public Data)

49. Drought Monitor – UNL (Public Data)
50. EveryCRSReport (Public Data)
51. Federal Register (Public Data)
52. FEMA Disaster Declarations (Public Data)
53. Forest Service Portal (Public Data)
54. General Service Administration (GSA) (Public Data)
55. Gov Info (Public Data)
56. Government Accountability Office (Public Data)
57. Johns Hopkins (Public Data)
58. NASS Quick Stats (Public Data)
59. NIFA Portal (Public Data)
60. NOAA National Hurricane Center (Public Data)
61. NWS Weather Alerts (Public Data)
62. Patents View – USPTO (Public Data)
63. RMA – Summary of Business & Cause Of Loss (Public Data)
64. USA Spending (Public Data)

1.3 Why is the information being collected, used, disseminated, or maintained?

Information is being collected, consolidated, synthesized and standardized into reliable source of data which will be used internally to significantly improve USDA's ability to make decisions related to management of the organization, strategic planning, operational effectiveness and delivery of programs and services to American citizens.

1.4 How is the information collected?

The information is collected in different ways including but not limited to data extractions through standardized Application Program Interfaces (API), submitted by source systems utilizing Secure File Transfer Protocol (SFTP), automated and manual batch extracts of xml, excel and csv files.

Data is a replication of the data from the source systems.

1.5 How will the information be checked for accuracy?

Information received will be stored in a staging area and data quality checks will be performed on the files received for integrity, accuracy and completeness of the data. The errors and inconsistencies will be logged in the error log tables and will be available for reviews. Depending on business requirements and use cases the data errors will be shared with the source systems for correction and re-processing.

1.6 What specific legal authorities, arrangements, and/or agreements defined the collection of information?

EDAPT team will be defining separate Interconnection Security Agreement (ISA) with each of the source systems except for publicly available federal data sources. The ISA documents methods of interfacing with the sources, data that is being shared, and information security techniques while transporting and storing the data.

The collection of information by source systems is governed by the existing system of record notice (SORN) and PIA. The EDAPT contains replicated data from these systems and does not directly collect new information about individuals beyond what is replicated from these sources.

1.7 Privacy Impact Analysis: Given the amount and type of data collected, discuss the privacy risks identified and how they were mitigated.

The access to data is restricted by strong access control methods including e-Authentication, PIV card integration and integration with USDA's ICAM. The data collected is encrypted both while in transit as well as at rest utilizing approved encryption techniques. The role-based access control will be implemented to enforce the separation of duties and least privileged access to the data. The products such as Cloudera Navigator and Cloudera Manager will be utilized to monitor and control access to the data as well as to maintain the audit of who accessed the data and when. All systems interacting with EDAPT are required to have appropriate security controls, this includes the Accenture Insights Platform (AIP) which is the hosting facility.

Section 2.0 Uses of the Information

The following questions are intended to clearly delineate the use of information and the accuracy of the data being used.

2.1 Describe all the uses of information.

The data collected will be standardized, cleansed and synthesized into consolidated data repositories to be utilized for analytical and decision-making purposes in the areas of HR, Finance, Payroll, IT systems, Properties and Mission Area Analytics etc. The data will be utilized by users to develop dashboards, reports and other data visualization to support the various business use cases for specific decision analysis. Data will also be made available to be queried (i.e. SQL) to a subset of users for self-service analytics. Data will be summarized and aggregated as needed basis and can be integrated with third party data such as geolocations etc. in future.

The SRA custom application collects non-PII budget data that is input by users of the application for transactional purposes.

Payment Schedules is a custom application that collects scenario and cost component data entry for transactional purposes.

The Informatica Enterprise Data Catalog and Axon store technical and business metadata to help give users a holistic view of metadata for inventory, organization, and analytical purposes. No source data is stored in these applications.

The Data Science Workbench is a 3-month pilot program where users access virtual environment to leverage a variety of pre-installed software tools, including SAS, ArcGIS, Python and RStudio. The virtual environment also offers processing power not normally available on a standard-issue government laptop, allowing users to create and test more efficient workflows.

IMAGES (Integrated Modeling and Geospatial Estimation System) provides a virtual environment of high-performance compute resources and advanced analytics tools to NASS data scientists, to enable them to enable them to serve the agricultural and rural data needs of the Nation with accurate, timely, and unbiased statistical information and services to the public. Virtual Machines are configured with many analytic tools including ArcGIS Pro, See5, ERDAS Imagine, R, Python, GDAL, PROJ, GEOS, and SAS.

DRUID (Data Repository and User Interface Design) provides infrastructure and tools to support the reimagined version of NASS QuickStats. NASS serves the basic agricultural and rural data needs of the Nation with accurate, timely, and unbiased statistical information and services to the public. NASS currently provides statistical information through a data dissemination tool called QuickStats. NASS intends to modernize this tool by adopting a cloud platform with current technologies for web portal design as well as implementing newer dissemination options such as data visualizations.

2.2 What types of tools are used to analyze data and what type of data may be produced?

Data is analyzed by a suite of statistical analysis, business intelligence and analytics software packages including Tableau, SAS, R, and Python etc. The data produced from these software suites include tabular data, graphical data, sensor data, image data, Geospatial data.

Custom web applications for SRA and Payment Schedules are used for transactional purposes.

The DRUID application is currently in development (deployment target: 09/2021) and will provide a Web UI and API. The web app will provide analytics (reports, stats, dashboards) and extract access via API. This is not a “data-entry” application.

Citrix virtual environment used for DSW and IMAGES to provide users with high-performance compute resources and advanced analytical tools to enable data scientists to perform statistical analysis and business intelligence.

Informatica EDC/Axon for analysis and governance of metadata.

2.3 If the system uses commercial or publicly available data please explain why and how it is used.

The system will utilize publicly available federal data listed in section 2.1 for Mission Area dashboards to include, but not limited to USA Spending system to track purchase of goods and services for USDA missions. The system will utilize National Hurricane Center, National Weather Service, FEMA, Department of Energy, and other weather and utility data sets to assist in assessing the impact of severe weather events on US persons, property, and goods as they relate to USDA and/or USDA programs. Data from Agricultural Marketing Service MyMarketNews data sources will be used for Cotton Quality Report. Using COVID data sets from Discovery Labs and Johns Hopkins for USDA COVID dashboards.

2.4 Privacy Impact Analysis: Describe any types of controls that may be in place to ensure that information is handled in accordance with the above described uses.

The system implements various security concepts in order to ensure that information is handled appropriately. This includes, but is not limited to, the concept of least privilege, separation of duties, logging, real time alerting, access escalation prevention and detection, and Rules of Behavior requirements. The system complies with NIST 800-53 controls requirements. This includes controls covering access control, risk management, audit and accountability, awareness and training, contingency planning, identification and authentication, system and information integrity, incident response, maintenance, media protection and more. The system security complies with USDA requirements to ensure that information is handled appropriately

Section 3.0 Retention

The following questions are intended to outline how long information will be retained after the initial collection.

3.1 How long is information retained?

The source systems to the EDAPT are responsible for minimum retention of data per published records retention schedule. As a reporting system the EDAPT relies on the source system for this retention. Additionally, the EDAPT will keep data for a maximum period according to temporary data retention requirements provided by USDA records management. The visualization constructs are viewed as working items and may be retained through their usefulness, it is not necessary to retain the constructs permanently, as one need only return to the original holdings and formulas to recreate.

3.2 Has the retention period been approved by the component records officer and the National Archives and Records Administration (NARA)?

The USDA Records Management group has reviewed the usage and purpose of the system and determined that the system is not a generator of new records for the purposes of record management as a creator of dashboards/reports utilizing a combination of source system data. EDAPT contents represents extractions from the Staff Offices holdings, and are considered temporary data/records, the Staff office are the legal holders of the source data/records and will determine retention of the holdings by file plan. Data residing in the EDAPT will be kept for a temporary period as defined by USDA records retention group.

3.3 Privacy Impact Analysis: Please discuss the risks associated with the length of time data is retained and how those risks are mitigated.

All data, whether new or old, is exposed to the same potential risk at any given moment. The access to data is restricted by strong access control methods including e-Authentication, PIV card integration and integration with USDA's ICAM. Access to system is controlled via role-based access control which will enforce separation of duties and limit access to the data while providing adequate access to the system components for administrative purposes. The data will be encrypted at all the time whether in motion or at rest. As mentioned in sections above the system will comply with NIST 800-53 control requirements for security of system and data within thus reducing the risk of unauthorized access and exposure of data.

Section 4.0 Internal Sharing and Disclosure

The following questions are intended to define the scope of sharing within the United States Department of Agriculture.

4.1 With which internal organization(s) is the information shared, what information is shared and for what purpose?

Information may be shared with OCIO, OCFO, OHRM, OPPM, OHSEC, OO Management, and USDA Mission Areas for analytical purposes through Tableau dashboards and reports. The information shared will include data from different business domains including Finance, HR, Payroll, Travel, Onboarding, Recruiting, Procurement, IT investments, Properties and Facilities, Facility access data (entry and exit), service request and IT application inventory, USDA Mission Area program and financial metrics, etc.

4.2 How is the information transmitted or disclosed?

The information will be transmitted to the USDA EDAPT via secure file transfer methods such as SFTP, ODBC, API, and Web Service. Once in the USDA EDAPT, data will be shared via a series of Tableau dashboards or via encrypted ODBC connections (i.e. SQL)

SRA and Payment Schedules custom applications collect scenario and cost component through data entry for transactional purposes.

NASS DRUID Custom Web App provides a public facing API for sharing data; this application does not contain PII.

4.3 Privacy Impact Analysis: Considering the extent of internal information sharing, discuss the privacy risks associated with the sharing and how they were mitigated.

Risk when sharing within USDA is considered low-moderate. Should data sharing include sources of the network, encryption protocols ensure PII is not inadvertently shared in an unencrypted format. Data is encrypted in motion and at rest. In addition, access to data is limited to only those persons with a need-to-know through the use of internal, granular governance process. Dissemination of information is governed by internal policy. Access to information is monitored, tracked, logged and audited using tools such as Cloudera Navigator and Cloudera Manager.

Section 5.0 External Sharing and Disclosure

The following questions are intended to define the content, scope, and authority for information sharing external to USDA which includes Federal, state and local government, and the private sector.

5.1 With which external organization(s) is the information shared, what information is shared, and for what purpose?

Public Tableau access will provide external users with read-only query access to publicly available information, does not contain PII, and will not have unfettered access to the EDAPT. Access to the EDAPT requires a privileged user account, GFE, and a PIV card. Nothing can be accessed on the EDAPT without all of these components.

NASS DRUID Custom Web App provides a public facing API for sharing data; this application does not contain PII.

5.2 Is the sharing of personally identifiable information outside the Department compatible with the original collection? If so, is it covered by an appropriate routine use in a SORN? If so, please describe. If not, please describe under what legal mechanism the program or system is allowed to share the personally identifiable information outside of USDA.

PII information is not shared outside of the USDA.

5.3 How is the information shared outside the Department and what security measures safeguard its transmission?

Not Applicable

5.4 Privacy Impact Analysis: Given the external sharing, explain the privacy risks identified and describe how they were mitigated.

Not Applicable

Section 6.0 Notice

The following questions are directed at notice to the individual of the scope of information collected, the right to consent to uses of said information, and the right to decline to provide information.

6.1 Does this system require a SORN and if so, please provide SORN name and URL.

The information in the system is derived from other source systems and it is not defined as a System of Records. EDAPT would leverage existing data source system SORN(s).

6.2 Was notice provided to the individual prior to collection of information?

The system does not directly collect data from individuals.

6.3 Do individuals have the opportunity and/or right to decline to provide information?

No because information is not gathered directly from individuals.

6.4 Do individuals have the right to consent to particular uses of the information? If so, how does the individual exercise the right?

Not Applicable

6.5 Privacy Impact Analysis: Describe how notice is provided to individuals, and how the risks associated with individuals being unaware of the collection are mitigated.

Not Applicable

Section 7.0 Access, Redress and Correction

The following questions are directed at an individual's ability to ensure the accuracy of the information collected about them.

7.1 What are the procedures that allow individuals to gain access to their information?

The system does not collect data directly from the individuals and is not a system of records and depends on source systems for the data. Hence the individuals are not allowed to access the data within the system.

7.2 What are the procedures for correcting inaccurate or erroneous information?

Since system depends on the data from source systems, it assumes that the data provided by source system is accurate. Any errors found are reported back to the source systems for correction.

7.3 How are individuals notified of the procedures for correcting their information?

Since system depends on the data from source systems, procedures to notify individuals will be coordinated by source systems.

7.4 If no formal redress is provided, what alternatives are available to the individual?

Since system depends on the data from source systems, redress will be provided by source systems.

7.5 Privacy Impact Analysis: Please discuss the privacy risks associated with the redress available to individuals and how those risks are mitigated.

Since redress will be provided by source systems, it assumes that the privacy risks associated with the redress will be mitigated by the source systems.

Section 8.0 Technical Access and Security

The following questions are intended to describe technical safeguards and security measures.

8.1 What procedures are in place to determine which users may access the system and are they documented?

Access to records in the system is limited to authorized personnel whose official duties require such access and is based upon the principle of least privilege. Data is protected through network two-factor authentication, unique usernames and complex passwords, database permissions, software controls, and encryption. The end users will have access to the data based on their affiliation to the specific departments and groups defined specific to those departments. Also, the access will be further restricted by role-based access control. The system administrators,

network administrators and database administrators will have broader access to system and databases to maintain system operations and performance but will not have access to the data.

8.2 Will Department contractors have access to the system?

Yes

8.3 Describe what privacy training is provided to users either generally or specifically relevant to the program or system?

The users will be provided annual security and privacy training as required by USDA policies.

8.4 Has Certification & Accreditation been completed for the system or systems supporting the program?

The C&A process has been completed for the system.

8.5 What auditing measures and technical safeguards are in place to prevent misuse of data?

The system has audit capabilities that allow the ability to perform incident response and investigation type activities. The system has real time alerts set up for known potential misuse for behaviors such as privilege escalation. The system complies with NIST 800-53 requirements that help prevent the misuse of data such as routine audits and log retention requirements. All information stored in this system is secured by utilizing database security technology and is resistant to tampering and circumvention by unauthorized users. Access to data by all authorized users will be monitored using both automated and manual controls. The information is accessed by authorized users on a "need-to-know" basis and intended systems usage basis. Every event accessing data, which includes but not limited to: SQL Queries, scripts, batch extractions etc, is logged by Cloudera processes. The Cloudera Navigator will allow cataloging of data and monitoring and managing of data access logs. It will also allow reports and charts to review the audit logs.

8.6 Privacy Impact Analysis: Given the sensitivity and scope of the information collected, as well as any information sharing conducted on the system, what privacy risks were identified and how do the security controls mitigate them?

The risk that personally identifiable information will be used inappropriately is mitigated by security training and by the use of audit mechanisms that log and monitor user activity.

Section 9.0 Technology

The following questions are directed at critically analyzing the selection process for any technologies utilized by the system, including system hardware and other technology.

9.1 What type of project is the program or system?

The system is a FedRamp certified GovCloud data lake solution which collects, standardizes and consolidates the data from various administrative systems across USDA into a central data repository to support data analysis, statistical and predictive analytics and decision making for OCIO office and various departments of USDA.

9.2 Does the project employ technology which may raise privacy concerns? If so please discuss their implementation.

No

Section 10.0 Third Party Websites/Applications

The following questions are directed at critically analyzing the privacy impact of using third party websites and/or applications.

10.1 Has the System Owner (SO) and/or Information Systems Security Manager (ISSM) reviewed Office of Management and Budget (OMB) memorandums M-10-22 "Guidance for Online Use of Web Measurement and Customization Technology" and M-10-23 "Guidance for Agency Use of Third-Party Websites and Applications"?

Yes

10.2 What is the specific purpose of the agency's use of 3rd party websites and/or applications?

OCIO will not use 3rd party websites/applications to store EDAPT data.

10.3 What personally identifiable information (PII) will become available through the agency's use of 3rd party websites and/or applications.

No PII is made available through third party website or application.

10.4 How will the PII that becomes available through the agency's use of 3rd party websites and/or applications be used?

Not Applicable

10.5 How will the PII that becomes available through the agency's use of 3rd party websites and/or applications be maintained and secured?

Not Applicable

10.6 Is the PII that becomes available through the agency's use of 3rd party websites and/or applications purged periodically?

Not Applicable

10.7 Who will have access to PII that becomes available through the agency's use of 3rd party websites and/or applications?

Not Applicable

10.8 With whom will the PII that becomes available through the agency's use of 3rd party websites and/or applications be shared - either internally or externally?

Not Applicable

10.9 Will the activities involving the PII that becomes available through the agency's use of 3rd party websites and/or applications require either the creation or modification of a system of records notice (SORN)?

Not Applicable

10.10 Does the system use web measurement and customization technology?

No

10.11 Does the system allow users to either decline to opt-in or decide to opt-out of all uses of web measurement and customization technology?

Not Applicable

10.12 Privacy Impact Analysis: Given the amount and type of PII that becomes available through the agency's use of 3rd party websites and/or applications, discuss the privacy risks identified and how they were mitigated.

No PII is made available to the third-party websites or applications.

Responsible Officials

THEODORE KAOUK
Digitally signed by THEODORE KAOUK
Date: 2021.06.15 09:01:34 -04'00'

June 15, 2021

Ted Kaouk — Date

System Owner – Chief Data Officer
Office of the Chief Information Officer
United States Department of Agriculture

LISA MCFERSON
Digitally signed by LISA MCFERSON
Date: 2021.06.15 13:27:32 -04'00'

June 15, 2021

Lisa McFerson — Date

Information Systems Security Manager (ISSM)
Office of Chief Information Officer (OCIO)
United States Department of Agriculture