Privacy Impact Assessment
Business Analytic Division (BAD)

Policy, E-Government and Fair Information Practices

- Version: 1.4
- Date: March 11, 2021
- Prepared for: USDA OCIO-Policy, E-Government and Fair Information Practices (PE&F)

USDA
United States Department of Agriculture

Privacy Impact Assessment for the

Business Analytic Division (BAD)

March 11, 2021

Contact Point

Jared Burnett
Information System Owner
(214) 767-7710

Reviewing Official

James Flickinger
Chief Information Security Officer, FPAC
United States Department of Agriculture
(816) 926-6010

Abstract

Name of the component and system: **BAD Data Warehouse (BAD)**

The BAD Data Warehouse system administered by the Center for Agribusiness Excellence (CAE) at Tarleton State University (TSU) is a data warehousing and data mining project for the United States Department of Agriculture (USDA) Risk Management Agency (RMA). The BAD Data Warehouse provides access to and maintains weather and crop loss data by way of a web application known as Hyper Dynamic Reporting Application (HyDRA). This Privacy Impact Assessment (PIA) is being conducted because of the vast amount of Personally Identifiable Information (PII) that is accessed through the BAD Data Warehouse.

A Privacy Threshold Analysis (PTA) was performed, indicating that a PIA must be completed. This PIA is being conducted to comply with the Federal Information Security Modernization Act of 2014 (FISMA) (44 U.S.C. §3551 to §3559) and the E-Government Act of 2002 (Pub. Law. 107-347, 44 U.S.C. §101).

Why the PIA is being conducted: To support federal law, regulations and policies.

Overview

System Name: BAD Data Warehouse (BAD)

System Description: The mission of USDA RMA Business Analytics Division (BAD) is to provide research, training, and resources for data warehousing and data mining of Farm Service Agency (FSA) and Risk Management Agency (RMA) data. BAD provides facilities and staff to make full use of data warehousing and data mining technologies for development of RMA products.

The primary tool used to securely provide information to BAD Data Warehouse users is through the Hyper Dynamic Reporting Application (HyDRA). HyDRA is an online resource that enables users to access weather data and numerous reports.

Authority to use this data is granted by RMA.

Legal Authority: This system is regulated by privacy laws, regulations and government requirements, including the Privacy Act (5 U.S.C. §552a); the E-Government Act of 2002 (Pub. Law. 107-347, 44 U.S.C. §101); the Paperwork Reduction Act of 1995 (44 U.S.C. §3501); the Federal Information Security Modernization Act of 2014 (FISMA) (44 U.S.C.

§3551 to §3559); OMB Memos M-03-22, M-10-22, M-10-23, M-16-24, and M-17-12; and OMB Circular A-130, Appendix I.

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?

RMA, FSA, and weather data is loaded and maintained in BAD databases. Data is transmitted from FSA and RMA, data is collected by the respective agencies and not from BAD. The following data is imported from RMA and FSA:

Producer name, address, phone numbers, tax ID, ID type, and entity type, as well as any associated farm numbers. Crop data includes state, county, crop, crop type, irrigation practice, intended use, share, field, and common land unit (CLU) which contains latitude and longitude, and etc. Policy data includes name, address, phone number, policy number, unit, Approved Insurance Provider (AIP), crop, insurance plan, coverage level, indemnity, liability, premium, fund placement, acres, agent name, agent social security number, adjuster name, adjuster social security number, and etc. The system does not utilize or store payment source information

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

The data imported from the source systems:
- RMA data is pulled from 165.221.67.8\PRODSQLB.55572
- FSA downloads are made from armmokcp3db29pv.usda.net
- CLU data is retrieved from: \\aiodcwa23fp20\FSASHARE\GIS.

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

The information is used to provide resources, training, and research for data warehousing and data mining of FSA and RMA data for the Federal Crop Insurance Program. The objective is to provide tools, information, and analysis proven effective in the private sector in reducing insurance waste, fraud, and abuse.

1.4 How is the information collected?

RMA Data Loads

Step One – Initial Data Load
1. The initial RMA data load into the BAD data warehouse was completed in June 2001.
2. Some data files consistently failed to load because of escape characters, e.g., backslash pipe (\|) or double backslash pipe (\\|) in the data. Those characters are removed, and the files are then loaded. Thus, identifying and correcting the files that do not load ensures the completeness of the data load.
3. Data was cleansed and standardized. Extract, Transform, and Load (ETL) tools and post cleansing scripts are used to clean up the data. For example, "Mr." will appear in some records with a period and in others without a period. Consistency changes are achieved by including/removing the period, thus eliminating the one-character difference that would be problematic in data analysis where unique individuals must be identified.
4. RMA data is maintained in separate annual databases by years. Data from all years is moved into a single production database within the BAD data warehouse. A validation script is run against the production database for the purpose of verifying conformation of the data to the RMA Manual 13.

Step Two – Periodic Data Updates
(Steps for weekly updates are consistent with the initial data load)
1. RMA data is loaded weekly, following the data update process (discussed later).
2. Based upon experience, there are some data files that consistently fail to load because of escape characters, e.g., backslash pipe (\|) or double backslash pipe (\\|) in the data. Those characters are removed, and the files are then loaded. Thus, identifying and correcting the files that do not load ensures the completeness of the data load.
3. Another aspect of the data update process is to cleanse and to standardize the data. ETL tools and post cleansing scripts are used to clean up the data. For example, "Mr." will appear in some records with a period and in others without a period. Consistency is achieved by including/removing the period, thus eliminating the one-character difference that would be evident in data analysis.
4. Although the RMA data is maintained in separate annual databases by years, it is moved to a single production database within the BAD data warehouse. A validation script is run against the production database for the purpose of verifying confirmation of the data to the RMA Manual 13.

FSA Data Loads
The initial FSA data load into the BAD data warehouse was completed in December 2009. FSA data is updated on a weekly schedule. The ETL and validation procedures applied to FSA data are similar to those used on RMA data except for the following cases:
1. Some records are being rejected due to the redundant commas (,,) in comma delimited source files. BAD identified and removed the redundant commas to ensure the completeness of the FSA data load.
2. Two sets of Microsoft SQL Server backup files are included in the source data. BAD created a workflow to restore the .bak files into a SQL Server database and to migrate the tables to the BAD data warehouse (DW). Validation logs of the restore procedure are used to verify that all data was successfully processed. Additional validations and edit checks are performed after the data integration to ensure completeness and accuracy of the final data.

3. FSA table row counts are verified to match either the original number of FSA records received or the record count in the SQL Server database.
4. The BAD DW does not have separate annual databases for FSA data.

For the row count test, electronic copies of the results are filed and include a date and time stamp from Teradata. Results of the validation scripts are in an electronic format and are available to the RMA project manager.

This report addresses database validity and reliability; therefore, the significant issue of product validity and reliability should also be addressed. Knowledge of the relevant federal regulations, industry standards, and RMA policy and procedure is essential for the BAD data warehouse specialist and research analyst to use in determining data validity and to ensure product integrity. The BAD DW product development and analysis includes a careful and comprehensive review of pertinent and applicable RMA rules and regulations and associated documents, e.g., Manual 13, common and crop policy provisions, loss adjustment procedures, and commodity handbooks as well as industry literature. The assessment and the results of the information are then discussed with applicable RMA specialist. It is the collective and cumulative knowledge and experience of BAD's personnel in understanding the RMA database and identifying data issues that ensure both data and product validity and reliability, and ultimately play an important aspect in quality control.

1.5 How will the information be checked for accuracy?

Software and Data Edit Checks
This involves the use of software checks and edits of data on the BAD data warehouse. The BAD data warehouse uses weekly testing by applying logical tests to electronic data files and hard copy reports of the data in the BAD data warehouse. The data is continually scrutinized. The following tests are applied to the data:
1. Record/row counts should match the original number of records/rows received from RMA.
2. BAD DW checks table structure to verify any changes in the number of attributes (added or deleted attributes). Post-row-count checks confirm that no records are lost in the update process.
3. BAD DW checks the relationship of one data element to another, e.g., by joining the crop unit line item table to the claim line item table.
4. Reasonableness check on RMA data for the following cases:
   a. Values outside of a designated range, i.e., amounts that are too small or too large as specified in the Manual 13
   b. Fields with null values
   c. Dates outside valid time frames or in an illogical progression
5. Elimination of duplicate records in tables.
6. Data totals of numeric data (row and column sum checks).
7. Any field showing large changes from one load to the next is investigated.
8. Data verification of the BAD database with the RMA source data by reinsurance year:
   a. Comparison of previous indemnity and row count with new indemnity and row count.
   b. Comparison of previous liability and row count with new liability and row count.

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

USDA/FSA-2 – Farm Records File (Automated).

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

Please refer to the BAD System Security plan. Any data collection risk are mitigated by limited access to data through logical access controls.

Section 2.0 Uses of the Information

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

2.1   Describe all the uses of information.

The information is used to provide resources, training, and research for data warehousing and data mining of FSA and RMA data for the Federal Crop Insurance Program. The objective is to provide tools, information, and analysis proven effective in the private sector in reducing insurance waste, fraud, and abuse.

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

Teradata and Oracle data bases are used to store, clean and produce usable information from raw data. HyDRA, the web application tool developed by the BAD DW is used to disseminate the information.

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

Publicly available and commercial weather data is used to validate claims due to weather related cause of loss.

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.

Only authorized users are allowed access to the information.

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?

Information will be retained for the life of the project then returned to RMA. The information received is used in the data mining and data warehousing for the agency. The data is maintained for historical analysis, thus is retained for the life of the project.

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

Yes, in accordance with USDA Directive DR 3080-001: Appendix A: Scheduling Records.

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

Retention period has no effect on the risk associated with PII. Access to the data is controlled through e-Authentication.

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?

Currently the information described previously is provided to FSA, RMA, and OIG within USDA. There have been limited OIG users that have access to the system, this was to obtain reports and transmit products and is a routine use.

4.2   How is the information transmitted or disclosed?

FSA and Approved Insurance Provider (AIP) HyDRA are accessible via caehydralink which resides on the RMA Extranet. An eAuthentication account as well as HyDRA authentication are required to gain access to this web URL.

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

Because all users with access to Hydra are authorized to view the information provided, risks are minimal.

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?**

Access to data stored in the BAD data warehouse is restricted to authorized users in RMA, FSA and AIPs. AIPs are only allowed access to data associated with their policies for the purpose of reporting.

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.**

N/A

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

N/A

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

N/A

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.

Business Analytics Division and the Center for Agribusiness Excellence do not collect the Personally Identifiable Information (PII) it is received from Farm Service Agency (FSA) and the Risk Management Agency (RMA). The use of the data is covered by the Federal Register / Vol. 77, No. 50/ Wednesday, March 14, 2012/ Notices. Department of Agriculture, Office of the Secretary, Privacy Act of 1974; Farm records File (Automated) System of Records. Agency: Department of Agriculture (USDA). Action: Notice of revision of Privacy Act system of records.
SORN short name FR Doc. 2012-6090 Filed 3-13-12; 8:45 AM

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

The BAD data warehouse does not collect information directly from individuals; information is obtained from FSA and RMA.

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

The BAD data warehouse does not collect information directly from individuals; information is obtained from FSA and RMA. The right to decline to provide information on the part of the individual is not applicable to this project.

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

The BAD data warehouse does not collect information directly from individuals; information is obtained from FSA and RMA. The right to decline to provide information on the part of the individual is not applicable to this project. Additionally, providing information on how an individual exercises the right to consent is the responsibility of the organization collecting the data.

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

N/A

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?**

Individuals are to contact their FSA county office, agent, or AIP to gain access to their data

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

Individuals are to contact their FSA county office, agent, or AIP for corrections to their data. In rare instances, data corrections can be sent to the RMA FOIA office.

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

RMA notifies the AIP, when necessary, about corrections made. The AIP in turn notifies the individual.

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

There is an appeals process involved to which the individual may apply.

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

Privacy risks are small but could include disclosure of the information during the appeals process. Access to the PII is restricted to authorized personnel. When data is transmitted from sources to CAE it is encrypted. When it is at rest in Teradata it is encrypted. When it is transmitted from in and out of Teradata it is not encrypted. This was discussed with USDA security in order to get permission to get rid of HP Sec (Voltage).

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?**

FCIC-586 are submitted to RMA for review, once approved the FCIC-586 is sent to BAD. The account is established in the BAD Web Application Administration Tool (ADMIN). Access to the ADMIN tool is restricted to designated BAD personnel.

8.2   Will Department contractors have access to the system?

Only BAD/CAE employees

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

BAD doesn't provide privacy training to AIP, FSA, or RMA users. Training is assigned by the agency and the AIPs are required to meet the requirements of the Standard Reinsurance Agreement (SRA).

USDA Information Security Awareness Training is required for FSA and RMA users. AIPs must meet SRA requirements.

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

Yes, the last BAD ATO was July 9, 2018.

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

Technological safeguards in place to protect data misuse include, but not limited to, user IDs, passwords, data encryption for secured FTP, intrusion detection and, file change detection.

8.6   <u>Privacy Impact Analysis</u>: 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?

Privacy risks include improper disclosure of data by employees. Controls to mitigate this risk include logical access controls and auditing.

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?

Data Warehouse and Data Mining.

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

No, the website is available to RMA users through the RMA network. AIP and FSA users are required to access the system utilizing eAuthentication login requirements.

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 Program Manager (ISSPM) 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?

N/A

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

N/A

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

N/A

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?

N/A

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

N/A

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

N/A

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?

N/A

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)?

N/A

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?

No

10.12 <u>Privacy Impact Analysis</u>: 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.

N/A

I have carefully assessed the Privacy Impact Assessment for the Business Analytic Division (BAD).

_____   _____
Jared Burnett   Date
BAD Information System Owner

_____   _____
Karl Hill   Date
Information Systems Security Program Manager

_____   _____
Amber Ross   Date
FPAC Privacy Officer