United States
Department of
Agriculture

USDA NASS
Data Lab Handbook

National
Agricultural
Statistics
Service

Data Lab and Data Access Group
Methodology Division

Revised
August 2024

Table of Contents

Introduction

Welcome to the National Agricultural Statistics Service (NASS) Data Lab. NASS values agricultural research and seeks to provide accurate and useful data in service to U.S. agriculture. The NASS Data Lab is managed by the Data Lab and Data Access Group (DLDAG) in the Methodology Division and is governed by the Data Access and Disclosure Review Board (DADRB).

This USDA NASS Data Lab Handbook (PSM-CS-02-Attachment-A-Handbook) is an official attachment of NASS Administrator Policy and Standards Memorandum PSM-CS-02 and is used to inform Special Sworn Agents (approved data users) accessing agricultural census and survey data for analytical research. **All data users are required to read and follow all policies and regulations described in this handbook – compliance is compulsory**. It consists of three sections that explain the policies, procedures, and support of the NASS Data Lab system and data access.

- Section 1 Regulations and Policies describes NASS principals and protocols that govern the use of NASS data. It explains the rules enforced to protect the confidentiality of the data and those who have provided it. Finally, it discusses NASS policy regarding the review of reports and articles prepared from the analysis performed.

- Section 2 Data Access and Disclosure Review describes the procedural function of the data access application process, clearance for Foreign Nationals, how to access the data, enclave requirements, and disclosure review.

- Section 3 Support describes the many data resources available through NASS.

Forward any questions to the DLDAG contact if clarification of NASS policies is needed.

1 Regulations and Policies

1.1 Special Sworn Status

Special Sworn Agents are authorized access to a specifically approved project with NASS dataset, variables, and specified objectives. The data were collected directly by NASS from farmers, ranchers, and agribusinesses and include statistics that quantify the agricultural practices of these operations. These data were collected under the authority of the Confidential Information Protection and Statistical Efficiency (CIPSEA) Act of 2018, Title III of Pub. L. No. 115-435, codified in 44 U.S.C. Ch. 35, which requires that all information be used <u>for statistical purposes only</u>. CIPSEA gives discretionary authority to NASS to swear in Special Sworn Agents to access NASS data. No individual has a right to the appointment nor is NASS obligated to appoint any individual. Data access can be provided to US-based academia from public or private universities, other USDA agencies, state government agencies with NASS cooperative agreements, and other Federal Statistical agencies for purposes that serve the public and contribute significantly to understanding the agricultural sector or the statistical procedures used by NASS to collect and summarize data. These data are not to be used for regulatory, enforcement, or investigative purposes.

Data enclave staff are authorized by NASS as official NASS Data Lab representatives. Each is trained to administer the regulations, policies, and operational procedures approved by NASS. Each assures that the data are accessed only by authorized users and all material leaving the enclave has been reviewed for disclosure risk.

1.2 Safeguarding Confidentiality

CIPSEA requires that no information which could be used to identify or closely approximate data for an individual farm, establishment, or enterprise be released to the public. NASS Special Sworn Agents must adhere to all policies, rules, and regulations of NASS, especially those governing the usage of confidential data. Data access is restricted to the dataset and project objectives explicitly defined in each approved agreement with NASS. These agreements include, but are not limited to, the following forms (descriptions for each are found in the sections below).

- ➢ *Standard Application Process Application* (SAP)
- ➢ *NASS Memorandum of Understanding* (MOU)
- ➢ *NASS ADM-042 Request to Access Unpublished Data (*ADM-042)
- ➢ *NASS ADM-043 Certification and Restrictions on Use of Unpublished Data* (ADM-043)
- ➢ *NASS ADM-044 User Attestation* (ADM-044)

In compliance with CIPSEA law, violations or breach of the confidentiality of any individual is subject to fines up to $250,000 and/or 5 years imprisonment. No information, which can be used to identify operators of individual farms, ranches, establishments, or enterprises, may be removed from the data enclave at any time, in any format. Datasets will reside within the data enclave only; they may not be copied or replicated to other mediums of any kind, including portable drives, phones, tablets, or laptops. Moreover, screen-sharing, snipping, or reproducing is strictly prohibited and subject to fines and penalties.

Special Sworn Agents may not discuss any record level data from individual farms, establishments, or enterprises with anyone outside of NASS unless they are approved to access the same dataset on the same project. Research must remain within the scope of the approved project.

Any data that could be used, directly or indirectly, to approximate the level of activity of any individual farm, ranch, establishment, or enterprise must be suppressed prior to release. NASS must review all materials requested for export from the data enclave for the possible disclosure risk of confidential information. Section 2.3 Disclosure Review of this document outlines output restrictions and necessary criteria for the disclosure review process.

NASS publishes several market-sensitive data series through the Agricultural Statistics Board. Release of these data is done in such a way to ensure everyone has access to the information at the same time and no one gains a competitive advantage. Requests for data that could compromise this objective will be denied.

1.3 Research or Analysis Products

Many NASS data users prepare reports of findings based on their research inside the enclave. Prior to publication, data users must include a citation on the source of the data and independent conclusions derived. NASS requests that researchers provide a copy of their finalized work for documentation.

All researchers must cite the National Agricultural Statistics Service as the source of the data used in the research. The analysis, interpretation, and conclusions are those of the author and are not subject to NASS review. Researcher results do not represent the views or opinions of NASS. The following is an example statement of source and disclaimer [*adjust specifications*, as needed] **that must be included** in any publication:

Summaries were derived using data collected in the [2022 Census of Agriculture] by the National Agricultural Statistics Service, United States Department of Agriculture. Any interpretations and conclusions derived from the data represent author viewpoints and are not necessarily those of NASS.

The following is an example of an availability statement for use with a publication submission:

Due to the NASS Confidentiality Pledge, the data underlying this article cannot be shared publicly. Secure access of NASS data may be obtained by agreement and sworn status only; restrictions apply.

2 Data Access and Disclosure Review

2.1 Preparing and Submitting a Proposal

Data access projects are granted primarily for statistical analysis seeking modeled results (inference, regression coefficients, standard errors, cluster, and the like) and not to reproduce summary tables of descriptive statistics. Summary statistics (e.g. variable means, totals, etc.) are

allowed, but only to the extent that they support modeling output. In this case, sample table shells must be provided in the SAP application.

NASS encourages users to limit the number of files they request for export from the enclave to the output necessary to produce the final report or paper. Restrictions are applied on the type and size of research results that can be released from the enclave. The output expected to be released are regressions and specified tables as found in an article in a peer-reviewed academic journal.

The SAP application sections *Methodology* and *Requested Output* are important in assessing both the substance of the proposal and the risk of disclosing confidential information. Applications must clearly indicate that the project will emphasize modeling output. Researchers who desire disclosure review of large volumes of tabular output must request a Special Tabulation rather than requesting access to record-level CIPSEA protected data (see Section 2.3).

Access to NASS data via a data enclave requires the approval by the Chair, Agricultural Statistics Board and the Data Access and Disclosure Review Board (DADRB). Depending on the type of data, researchers must request access using the SAP online application process, or in special circumstances, an ADM-042. The research team must be associated with a U.S. institution, agency, or company and all researchers must access the data from within the United States.

Applications are evaluated on a project-by-project basis; that is, an MOU or ADM-042 is required for each project, with specifications. Census of Agriculture variables must be submitted at the time of the request and are instrumental in the decision-making process. Although amending the variable list is possible, researchers are advised to submit a comprehensive list from the onset. Data ingestion fees apply. Access is granted based upon many factors, including the fitness-for-use of the data for the stated objectives and minimally selected variables that explicitly target those objectives.

Applicants must include all components of the SAP and must submit a list of variables for Census of Agriculture data via NASS-provided spreadsheet. The geographic level of aggregation to be shown in the output must be included in the project requested output question.

A request can be made online through the SAP process and questions may be sent to SM.NASS.Data.Lab@usda.gov.

2.1.1   Standard Application Process

The Standard Application Process, or SAP, is the primary way to apply for access to protected data from all 16 federal statistical agencies and units for evidence-building. For NASS, the SAP application process results in an MOU agreement ready for review and approval consideration.

The MOU is an agreement between NASS, the institution/agency, and the researchers. Requests by academic institutions must be made by a professor/advisor project lead; students, who are not official agents of the institution, are not permitted to be project leads. NASS strongly encourages advisors to have an enclave account if they would like to see the data or output prior to the final disclosure review.

The SAP website includes a metadata catalog of protected data assets, or "restricted data," available across each of the federal statistical agencies. Through this site, researchers can determine if restricted data are appropriate for specific statistical research objectives and apply directly online at researchdatagov.org.

2.1.2   ADM-042 Request to Access Unpublished Data

For special circumstances, the ADM-042 is used to request access to record-level data. This form is reserved for parties who already have an external project agreement with NASS for survey design, testing, and/or collection.

2.1.3   Amending an Approved Project

Over the course of a project there may be reason to amend an approved application (e.g., changes in the individuals on the project, updated methodologies, etc.), this can be done by amendment. When reviewing an amendment request and its justification, the DADRB will consider if the proposed changes are outside the scope of the project and require a new application. The request must be submitted to the Data Lab manager by the project lead and accompanied by compelling justification that the changes are essential. Note that the project lead on any given approved agreement cannot be changed; in this case, a new application is required.

2.2   Data Enclave

The data enclave is a platform that provides secure remote data access to approved researchers. It is owned and operated by a private vendor and is fee-based; all enclave accounts must be paid by the data user. If the data user has established an external project agreement with NASS, the enclave fees can be incorporated into the cost estimate. If not, they must be paid directly to the vendor. The platform provides researchers with remote access to approved datasets for analysis within the enclave via a secure internet connection using the researcher's computer and web browser from an approved U.S. site location. Researchers will have access to a suite of analytical software to use within the data enclave. For the user, the data enclave appears as a cloud-based Microsoft Windows environment.

Per project, the enclave costs involve a user-setup fee and annual project workspace fees. There are configuration options based on researcher needs including CPU, Memory, Storage, Compute Hours, Data Ingestions, and Export Requests. Additional services beyond configuration specifications may be purchased by invoice. Invoicing is handled directly by the vendor. Each enclave account is non-transferrable and cannot be shared. Multiple researchers approved on a single project will share the same virtual space, have access to the same approved project folders, but must have their own account in the enclave. The data enclave vendor will discuss pricing options with researchers before onboarding.

2.2.1   Requirements

Data enclave managers will provide complete and comprehensive training, onboarding, invoicing, and a full list of applications available from within the enclave. To access NASS data via the data

enclave, the following is required:

2.2.1.1 Authorization from NASS to access unpublished data via an approved MOU or ADM-042.

2.2.1.2 A computer with internet access and a web browser. The computer must be limited to a fixed site and IP address.

2.2.1.3 An approved virtual site inspection of the access point, or place from which the data enclave will be used, by NASS personnel. Access points must be within the United States and comply with the following:
Free from "high traffic"
Isolated from public areas within the building
Accessible by authorized individuals only
Located inside a lockable room
Monitor is not visible from windows or doors

2.2.1.4 Completion of vendor-provided data enclave training, onboarding, and invoicing.

2.2.2 Security Protocols

All data users with approved access to unpublished data will be required to accomplish all NASS Security Protocols. This will include confidentiality training, compliance with the stipulations outlined in the *USDA NASS Data Lab Handbook,* and submitted signed *NASS ADM-043* and *NASS ADM-044* forms – digitally, otherwise witnessed by a NASS employee. These documents attest to the understanding of the laws governing confidentiality of NASS data and the penalties of violating them. The ADM-043 contains pertinent excerpts of these laws in more detail. Security protocols are required annually and by project.

Foreign national researchers are permitted to access NASS data after an additional mandatory clearance process is conducted by the Department of Homeland Security. The process begins after project approval and can take 2-3 months after documents and fingerprints are submitted; clearance for Green Card holders takes significantly less time. If the foreign national is the project lead, no other members may access the data until the clearance is granted. If the foreign national is a member of the team, the project lead and other approved researchers may begin while the clearance process is underway. More information on this process can be obtained by emailing inquiries to the Data Lab and Data Access Group (SM.NASS.Data.Lab@usda.gov).

2.2.3 Process and Procedures

After the MOU or ADM-042 has been approved and all security protocols have been completed, a NASS representative will contact the vendor to request onboarding to the data enclave. An enclave manager will arrange to provide full instructions on training, account options, billing, and onboarding to the data enclave. Enclave accounts are individually obtained; they are separate and nontransferable.

The data enclave contains controls that prevent users from printing and transferring files to and from the environment via USB, e-mail, screen captures, or otherwise. The signed confidentiality documents strictly prohibit snipping, photographing, screen-sharing, replicating, or any such capturing or transferring of the data – even with other members of the approved project. Without a separate data enclave account, project members are authorized to collaborate through planning, oversight, and dialogue only. Researchers may not conduct any research or analysis with the approved data outside of the enclave or away from the approved data enclave access point.

All files intended for removal from the data enclave must first be reviewed by NASS personnel for disclosure concerns. Only files that have been reviewed by NASS and approved for removal may be removed from the data enclave. Requests for export are made in the data enclave environment; training will be provided by the data enclave manager, and additional criteria can be found in Section 2.3 Disclosure Review.

2.2.4 Access Point Location

In general, the access point for accessing the enclave must be from the professional offices of the institution – university, state department, etc. with the associated IP address. The approved location is not mobile; the site inspection, which must be cleared by the NASS security officer, is for a single chosen location only. Under special circumstances, a home office may be considered for approval.

2.2.5 Project Workspace Closeout

NASS limits data access agreements to three (3) years. Researchers must plan accordingly to accomplish the project goals in this timeframe. At the end of the access agreement, the workspace will be deleted.

The enclave vendor typically invoices for one (1) year enclave access. At the end of the enclave contract period, enclave users will receive a 60-day notification of account expiration in order to complete their work or renew the enclave contract. NASS requires at least 30 days for export request of output or statistical code. If the enclave contract is not renewed (up to 3 years), the workspace will be deleted.

2.3 Disclosure Review

All research results requested to be released from the data enclave must undergo a disclosure review conducted by NASS. This requirement is in accordance with the NASS obligation to protect the confidentiality of an individual farmer, rancher, or producer.

When the research has concluded, the disclosure review of the research results is conducted as a unit; that is, all files are reviewed together. Since the data are often correlated, multiple export requests for one project present increased disclosure risk and often increases suppressions. In addition to the current disclosure review, the NASS analyst must review all past export requests from other projects to ensure there is no overlap that results in disclosure concerns. Currently, the review process is done manually; reviewers read through statistical programming code and all

output files – first individually, then against each other. For this and other reasons, large tabulations and incremental or partial submissions are strongly discouraged.

NASS requires users to limit the number of files they request to export to the necessary output required to produce a final report or paper. Restrictions apply (see Section 2.1).

Revise and resubmit for peer-reviewed journal articles will be considered, if the previously approved variables are still available. Potential updates to already released research results requires an active access agreement, enclave account, and the detailed specifications including the published material.

2.3.1 Disclosure Methodology

NASS employs disclosure limitation methodology commonly used by many Federal statistical agencies. Disclosure analysis methods use one of two criteria to determine whether a cell presents a disclosure risk.

The first criterion is a threshold rule, where a minimum number of operations must produce the item before a total can be released. For example, if only two farms produce milk in a county, releasing the total milk production allows the two farmers to deduce the production of the other. NASS uses the same threshold rule for all disclosure reviews. Each summarized estimate must be computed from a minimum of five (5) unweighted* observations. This means, anytime there are less than five unweighted operations the entire table will be suppressed. Consider removing or collapsing cells that do not meet this threshold before submitting a disclosure review.

*Note: Any output intended for publishing must be weighted.

The second criterion is a dominance rule; NASS uses different dominance rules in different circumstances.

The (n,k) rule invokes a suppression when the top *n* producers account for *k* percent or more of the estimated total. For example, a (2,80) rule will suppress a cell when the top 2 producers represent 80 percent or more of that cell total.

The p-percent rule requires sufficient protection so that the largest producer value cannot be approximated to within a range of p-percent. For example, a 20-percent rule will suppress a cell if revealing that total allows someone to estimate the top producer value to within plus or minus 20 percent.

Federal statistical agencies do not publicly disclose the actual values of n, k, or p, as revealing them compromises the protection.

2.3.2 Suppressions

Cells that represent disclosure risk are defined as *primary cells*, which are always suppressed, and are called *primary suppressions*.

In many instances, a primary suppression requires another suppression to maintain protection. This is due to the additive relationships prevalent in tabular summaries; this is one reason, in part, the release of tabulations from the enclave is restricted. For example, a suppressed value that is one element of a total can be deduced by simple subtraction. If A + B + C= D, suppressing C alone gives the value of C no protection, as its value can easily be obtained by solving the equation for C.

Thus, additional suppressions are needed to ensure confidentiality. These are called *complementary suppressions*. In some instances, two primary suppressions can serve to protect one another. However, in general, selecting complementary suppressions is much more difficult and time consuming.

If NASS finds values that require suppression, the NASS analyst may apply the suppression or require the researcher to adjust and resubmit the results.

Examples of this action might be:

> NASS may suppress individual table cells by explicitly replacing the cell value with an indicator identifying a suppression. NASS uses a (D), in place of the actual number in the affected cells, to indicate the value is withheld to avoid disclosure of an individual operation.

> NASS may request the researcher to aggregate the data at a higher level in order to reduce the number of suppressed cells.

During every disclosure review NASS must review the results to verify fitness-for-use standards are maintained. Additionally, NASS must consider the disclosure results from all other publications and tabulations of the approved dataset(s). NASS may require additional suppressions be applied to ensure consistency and confidentiality.

2.3.3 Implicit Samples

Most projects in the Data Lab create implicit samples that often get overlooked by researchers. Data Lab researchers, NASS analysts, and members of the Data Access Disclosure Review Board are required to consider the implications of the approved research project including any implicit samples created during the course of the research. Implicit samples can be difficult to conceptualize. As described by the Federal Statistical Research Data Center Disclosure Avoidance Handbook: "an implicit sample is a sample that can be identified by looking at the differences between explicitly defined populations, samples, subsamples, and geographies." If implicit samples are not considered during the course of the research project and accounted for during a disclosure review the results can pose a disclosure risk to the implicit sample population. Researchers are required to identify all implicit samples relevant to their research and may be required to submit disclosure review documentation (See Section 2.3.5) for the implicit sample. In some cases, if the results on the explicit sample do not provide adequate protection for the implicit sample the results from the explicit sample may not pass disclosure review and will be suppressed.

Some examples of implicit samples are:

1. A research project aims to analyze cattle farms in the state of Wisconsin. The explicit sample in this case is cattle farms in Wisconsin. An implicit sample is non-cattle farms in the state of Wisconsin. Another implicit sample is non-cattle farms in the United States. Another implicit sample is cattle farms not located in Wisconsin.

2. A research project aims to analyze farms with organic sales which is considered the explicit sample. The implicit sample is farms without organic sales.

3. A research project aims to analyze government payments received by farms with at least one black producer. An implicit sample is farms who receive government payments who do not have at least one black producer. Another implicit sample is farms who have at least one black producer but do not receive government payments.

2.3.4 Disclosure Rules

Data access projects should be used primarily for statistical analysis seeking modeled results (inference, regression, cluster, etc.) and not to reproduce summary tables or descriptive statistics. The disclosure review process for NASS-created Special Tabulations is automated. Enclave results, however, are reviewed manually and therefore require more stringent disclosure rules.

The following rules must be observed when submitting requests for disclosure review for all results intended for removal from the enclave. These procedures are designed to expedite and simplify the process. To avoid unnecessary delays, review the following information about disclosure prior to producing summary statistics or aggregates.

2.3.4.1 Output using intervals must mirror the NASS publication from which the analysis is derived. For example, if the publication from which the analysis is derived includes acreage breakouts of 1-9 acres, 10-49 acres, 50 to 179 acres, 180 to 499 acres, 500 to 999 acres, 1,000-1,999 acres, and 2,000 acres or more, the desired output must use the same acreage intervals.

2.3.4.2 Every aggregated cell (modeled or summarized) in the output (counts or values) must include a minimum of five (5) unweighted[*] contributing observations.

2.3.4.3 If output includes demographics, there must also be a minimum of 30 weighted contributing observations reporting that single demographic at the geographic level of aggregation.

2.3.4.4 Duplicative output will not be reviewed – e.g. descriptive statistics in table and report forms.

2.3.4.5 Volume of summary (aggregate) statistical output must be parsimonious. Data Lab data exports are not meant to allow mass export of data. The Data Lab reserves the right to deny data export due to overly inclusive data exports.

2.3.4.6 All summary statistics must use the record-level survey or census weights that NASS provides with each data set. For modeling purposes, NASS strongly encourages the use of weighted counts or values for robust outcomes. Any output intended for publishing must be weighted.

2.3.4.7 All summary statistics must be rounded to four significant digits, base 10. Significant digits are those that are neither leading nor trailing zeros. For example, the following estimates are presented to 4 significant digits:

> 1,234,000
> 1,234
> 1.234
> 0.001234

2.3.4.8 Unlike other summary statistics, ratios, proportions and percents need to be rounded before their construction. Numerators and denominators should be rounded using the following rule:

- Values less than 15 cannot be used
- Values between 15 - 99 should be rounded to the nearest 10
- Values between 100 – 999 should be rounded to the nearest 50
- Values between 1,000 – 9,999 should be rounded to the nearest 100
- Values between 10,000 – 99,999 should be rounded to the nearest 500
- Values between 100,000 – 999,999 should be rounded to the nearest 1,000
- Values 1,000,000 or more should be rounded to four significant digits described in Section 2.3.4.7

Example Table:

	Before Rounding Applied	Rounding Applied for Disclosure Review
Male Producers	3265	3300
Female Producers	920	900
Total Producers	4185	4200
Ratio of Female Producers	920/4185 = 0.22	900/4200 = 0.21

2.3.4.9 Data points will not be authorized for removal from the data enclave. Visualizations and other graphs must be curvilinear and may not represent actual data points or maximum/minimum values.

2.3.4.10 Graphs and charts using bins must ensure they are wide enough to provide reasonable uncertainty as to the values of those records in the bin and must contain a minimum of five (5) unweighted[*] contributing observations. Any graph or chart data point, scale, or axis must be rounded to the 4th significant digit.

2.3.4.11 Histograms will not pass the disclosure review process. Rather, users should use kernel-density plots. Kernels with infinite support are preferred.

2.3.4.12 The Data Lab allows for the export of pseudo-percentiles rather than percentiles. A pseudo-percentile is constructed by taking the average value of observations around the actual percentile, with a minimum of 5 observations from both sides. For example, a pseudo-percentile for the median would be the average of the 5 observations just below the actual median, the median itself, and the 5 observations just above the median. This ensures that at least 11 observations are required to report a percentile. Each pseudo-percentile must contain individual observations. No pseudo-percentile can use observations that are also used to calculate another pseudo-percentile. No percentile reported can equal the value of the largest contributor (the max value) or the smallest contributor (the min value).

2.3.4.13 For model output, such as linear or nonlinear regression models, disclosure risk is less. However, NASS requires a minimum of ten (10) degrees of freedom (i.e., the number of unweighted[*] contributing observations minus the number of variables used in the model must be greater or equal to 10) and will review the model output to check for disclosure failures.

2.3.4.14 When the use of unpublished intervals is approved, the interval range must be sufficiently wide. As a rule, if 120% of the midpoint of the range exceeds the upper bound of the range than the interval is too narrow. Two examples are provided for clarity:

Range	= $500 - $1000	Range	= $100 - $110
Midpoint	= $750	Midpoint	= $105
Threshold	= Midpoint * 1.20	Threshold	= Midpoint * 1.20
	= $750 * 1.20		= $105 * 1.20
	= $900		= $126

Therefore, since max of range ($1,000) is greater than the threshold ($900), the range is wide enough.

Therefore, since the max of the range ($110) is less than the threshold ($126), the range is not wide enough and must be redefined or suppressed.

2.3.4.15 The analysis and summary results in an editable format:
MS Excel (.csv/.xlsx), or MS Word (.docx or other text file) is accepted. HTML, .pdf, and non-text Stata output are not accepted.

2.3.4.16 Since all output are reviewed as a unit, if any output does not comply, the entire submission must be rejected.

*Note: Any output intended for publishing must be weighted.

2.3.5 Submitting a Request for Disclosure Review

As described in the previous section, NASS must perform a disclosure review of all output produced from the data. The entire set of analysis for an approved data access agreement is reviewed at the same time. Incremental or partial submissions are strongly discouraged and may be rejected. NASS encourages users to limit the number of files they request for export from the enclave to the output necessary to produce the final report or paper. Restrictions are applied on the type and size of research results that can be released from the enclave. The output expected to be released are regressions and specified tables as found in an article in a peer-reviewed academic journal.

Submit all requests for disclosure review through the data enclave managers. The enclave manager will conduct an initial review and will forward it to the Data Lab and Data Access Group for full disclosure review. The review time depends on the complexity of the output, the number of reviews currently in progress, and the completeness of the submission. NASS requires at least 30 days for this process and recommends allowing for such in the project Timeline, which is requested in the SAP application.

The disclosure review submission must include documentation with the following information:

2.3.5.1 The contact information (email/phone) of the Sworn Special Data Agent who prepared the results for disclosure review.

2.3.5.2 A description of each output file, including modeling techniques, data used, and explanation of indication.

2.3.5.3 A complete list of data sources used in the analysis and summary including information on which results are derived using user-provided data in combination with NASS provided data

2.3.5.4 A complete list of variables used, with descriptions – including variables created for research purposes.

2.3.5.5 Details on any data or variable modifications made – data subset may be required.

2.3.5.6 Specific geo-political areas being investigated or summarized – state, county, etc.

2.3.5.7 Each summarized tabulated data cell in the disclosure review submission must be accompanied by the following in the same file (see Appendix A):

- The total weighted and unweighted number of contributing observations
- The unweighted value of each of the top two largest contributing records

2.3.5.8 If rounding was necessary for disclosure review (See Sections 2.3.4.7 and 2.3.4.8) the researcher must provide the results prior to rounding and those after rounding is applied.

2.3.5.9 If the disclosure review includes percentiles the researcher must provide the true percentile which will not be eligible for export and the pseudo-percentile which will be eligible for export.

2.3.5.10 The statistical programming code used to create the analyses, with appropriate documentation. That is, every code file should have a header describing the contents of the file, including a summary of the data manipulation that takes place in the file. HTML or compiled code files are not acceptable.

2.3.5.11 Visualizations must be accompanied by underlying summarized data.

2.3.5.12 A completed *NASS Export Request Checklist*, acknowledging compliance with these and all stipulations conveyed herein regarding the use of NASS data.

NASS reserves the right to request additional information not included in the above list to conduct a disclosure review. If additional information is needed a list of requirements will be provided to the researcher. Failure to comply with disclosure rules or provide the information listed above will extend the disclosure review time as re-submissions will be necessary.

3 Support

NASS complies with the timeframe for approval allowed by the SAP. It is recommended that researchers allow several weeks from the time the request is submitted to when access to the data is granted. This provides NASS adequate time to vet the request, present it for approval, and prepare the data set for transfer. This allows for production demands and the enclave onboarding requirements.

NASS highly recommends that researchers are aware of their institution/agency process for outside invoicing. Data access in the enclave is contingent upon finalized fee payment, and this process can increase the project timeline significantly.

Once access is approved, the data will be assigned for preparation and transfer to the enclave. This process time can vary by data source and current production demands. Changes to the agreement for data or researchers can be requested via amendment. A request to change the project lead necessitates a new project application and agreement. The data will not contain personal identifiers such as name, address, zip codes, latitude, or longitude. The default format for delivered datasets

is comma-separated values (.csv), which can be easily imported into statistical packages. Inquire if an accommodation is needed.

If the approved project cited additional non-NASS data sources required to accomplish the objectives, consult with the Data Lab manager for upload. Additional support documentation, including survey or census questionnaires or variable description lists, are available through the SAP.

Each research team will have a Data Lab manager, who is aware of NASS data products and specializes in the operation of the data enclave. The role of these individuals is to help you get started with your research by assisting through the approval process, conducting all security protocols, transmitting the dataset, and arranging for disclosure review.

3.1 Data Resources

Agricultural Resource Management Survey (ARMS) data collection are conducted in collaboration with USDA Economic Research Service (ERS). All requests for ARMS data are made through the SAP and any questions may be directed to the ERS ARMS Team at ARMSTeam@usda.gov

A few of the many online resources for NASS data are available at the following sites:

ARMS Questionnaires and Manuals
Census of Agriculture 2022
Guide to Products and Services
Quick Stats (data downloads)
ResearchDataGov (SAP metadata and application)
Surveys and Programs

Appendix A – Example Table of Tabular Data

Geography	Variable	Estimate	Weighted N	Unweighted n	Largest Observation (unweighted)	2nd Largest Observation (unweighted)
US	TVP	####	####	####	####	####
Alabama	TVP	####	####	####	####	####
Alaska	TVP	####	####	####	####	####
Arkansas	TVP	####	####	####	####	####
…						