



United States
Department of
Agriculture



National
Agricultural
Statistics
Service

Handbook for Special Sworn Data Users of NASS's Data Lab

**Data Lab
Census Data Section
Summary, Estimation, and Disclosure Branch
Methodology Division**

Revised
June 2017

Table of Contents

Contents

Introduction.....	
I. Regulations and Policies.....	1
Special Sworn Status	1
Data Lab Rules and Procedures	1
Safeguarding Confidential NASS Data	2
Safeguarding NASS Systems and Hardware	3
Research or Analysis Products.....	3
II. Data Lab Operations.....	4
Locations.....	4
Access and Confidentiality Certification	4
Primary Contact	4
Data Support	5
Disclosure Review	5
III. Computing Environment.....	7
Workstation Environment.....	7
Client Hardware	8
Operating System.....	8
Application Software	8
Operation of the Lab	8
Computing Support.....	9
IV. RFO Role and Responsibilities.....	9

Introduction

Welcome to the National Agricultural Statistics Service's (NASS's) Data Lab. We hope your research or analysis goes smoothly and is successful. We provide this handbook to all Special Sworn Data Users who are affiliated with NASS for the purpose of accessing agricultural data for research or analysis. This handbook consists of four sections that explain the policies and usage of the Data Lab.

- Section I describes NASS's policies and regulations that govern the operation of the Data Lab. It also explains the rules you must follow to protect the confidentiality of the data in the lab. Finally, it discusses NASS policy regarding the review of reports and articles prepared from the analysis performed in the lab. We ask that you read this section very thoroughly so you fully understand our policies, particularly those concerning confidentiality.
- Section II describes the operation of NASS Data Labs including requesting access, the support available to facilitate your research or analysis, and disclosure review.
- Section III describes the computing environment at the NASS Data Lab.
- Section IV describes the roles and responsibilities of the NASS Regional Field Office.

You are required to follow all policies and regulations described in this handbook. If you have any questions, please ask your lab contact or the appointed agent so there will be no confusion or misunderstandings about our policies. We welcome your comments on the content of this handbook or the policies set forth in it. Please submit any comments to this handbook or our policies to the lab staff. We will update this handbook, when necessary, and distribute it to lab users.

I. Regulations and Policies

Special Sworn Status

Special Sworn Data Users are authorized access to a designated NASS dataset. These data were collected directly by NASS from farmers, ranchers, and agribusinesses and they include statistics that quantify the agricultural practices of these operations. These data were collected under the authority of Title 7, Section 2204g, United States Code, which requires that all information be used only for statistical purposes. The Confidential Information Protection and Statistical Efficiency Act (CIPSEA) gives discretionary authority to swear in Special Sworn Data Users to NASS. No individual has a right to the appointment nor is NASS obligated to appoint a person. Data access can be provided to academia from public universities, other USDA agencies, state government agencies with NASS cooperative agreements, and other Federal Statistical agencies for purposes that serve the general public and contribute significantly to understanding the agricultural sector or of the statistical procedures used by NASS to collect and summarize data. These data are not to be used for regulatory, enforcement, or investigative purposes.

Any Data Lab set up in a non-NASS facility will be supervised by an approved NASS representative. The NASS representative will be trained by NASS to administer the policies, regulations, and operations of the lab. NASS Representatives assure the lab is accessed only by approved users and all material leaving the lab has been reviewed for disclosure.

Data Lab Rules and Procedures

A designated workstation(s) is provided for Special Sworn Data Users to conduct research or analysis. You MAY NOT conduct any of your research or analysis with the data away from this workstation. There are no permanent NASS employees assigned to these workstations. A lab contact will be assigned to help you with access and any questions you may have.

Upon arrival each day, you must sign in and sign-out on the lab log. The log is used to verify users are authorized and the usage data helps with performance reports, future planning, budgeting, and audits.

ONLY Special Sworn Data Users are authorized to be in the NASS Data Lab area. Any unauthorized visitors must be reported immediately to the lab staff, agent, or their supervisor.

Telephones are made available to you in NASS offices for business purposes only. You should also note that the open configuration of the Washington Data Lab makes lengthy and loud telephone conversations distracting to others using the lab. Please keep your telephone calls short and your voice volume down. You may not discuss any of the data contents with anyone over the phone. Mobile phones are not permitted in the workstation area of the lab.

Safeguarding Confidential NASS Data

Title 7, Section 2276 and CIPSEA require that no information that could be used to identify or closely approximate data for an individual farm, establishment, or enterprise be released to the public. You, as a NASS Special Sworn Data User, must adhere to ALL policies, rules, and regulations of NASS, especially those governing the usage of confidential data. Your access is restricted to the dataset and project explicitly defined in your agreement (ADM-042 and ADM-043) with NASS. Violation of the confidentiality of an individual's data is subject to fine and/or imprisonment.

No information, which could be used to identify the operators of individual farms, ranches, establishments, or enterprises may be removed from the Data Lab. Any data that could be used to directly or indirectly approximate the level of activity of any farm, ranch, establishment, or enterprise must also be suppressed. NASS must review all materials you wish to remove from the lab site for possible disclosure of confidential information including files, flash drives, printouts, tabulations, and notes. A lab contact will be assigned by the Regional Field Office, approved NASS representative, or Headquarters to arrange necessary disclosure review of all output you produce before it can be removed from the Data Lab. The next section discusses the procedures pertaining to the disclosure review. Our goal is to complete the disclosure review within 30 days, however this depends on the size and complexity of the output. Please review section II, Data Lab Operations, in this document about how to output your data for the disclosure review.

No microdata (data for individual farms, establishments, or enterprises) may be removed from the Data Lab in any format. Datasets reside on the Data Lab computer only and may not be copied to other mediums including portable drives, phones, tablets or laptops.

As a Special Sworn Data User, you may not mention or discuss any data from individual farms, establishments, or enterprises to anyone outside NASS unless they have approved access to the same dataset. Other Data Lab users may not have authority to the same dataset you do.

Any documents to be discarded that contain confidential data must be securely destroyed. Labs are equipped with paper shredders for your convenience. Consult your lab contact for local disposal procedures.

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

Safeguarding NASS Systems and Hardware

A standalone computer will be set up for your use. It will be configured with commonly used software products including SAS, R, and Microsoft Office. If special software is needed, your lab contact will work with you to find an acceptable solution. The computer will not have a communications gateway to the outside.

Any programs, electronic reference materials, electronic products, and data to be brought into the Data Lab by you must be cleared through lab staff or the NASS representative. Pirated software cannot be loaded to any Federal computer. Each cleared file or program, such as SAS code, will be marked and documented that it has been cleared. Our purpose is to ensure that no virus or contamination is inadvertently introduced into the NASS system. Any such materials on hand without clearance by the lab staff will be immediately confiscated. Only NASS lab staff can copy files, including SAS programs onto the computers.

Research or Analysis Products

Many Data Lab clients prepare reports of their findings based on their use of the lab. We ask you to provide a copy of the draft report to your lab contact regardless of whether it is for public distribution or internal use. NASS will review the report as a final check on disclosure and to be sure all descriptions of NASS procedures and methodology are accurate. We will also check tables and charts to ensure header and stub labels are correct. NASS collects data using sound designs and accepted statistical practice to produce quality datasets and we stand behind our data. Hopefully, NASS's reputation contributes to the credibility of your research.

We ask that you cite the National Agricultural Statistics Service as the source of any statistics you calculated using NASS data. A statement of source should read like the following example:

Summaries were derived using data collected in the 2012 Census of Agriculture by the National Agricultural Statistics Service, United States Department of Agriculture.

The analysis, interpretation, and conclusions are those of the author and not subject to NASS review. They must not appear to represent NASS. Therefore, a disclaimer statement must also be included, for example:

Any interpretations and conclusions derived from the data represent the author's views and not necessarily those of NASS.

II. Data Lab Operations

Locations

Washington, D.C. - NASS operates a full time Data Lab at its Washington D.C. headquarters in the South Agriculture Building. The Data Lab is part of the Census Data Section of the Methodology Division. This Data Lab has full time technical staff.

NASS Regional Field Offices - NASS will set up a Data Lab in any of its 12 Regional Field Offices if there is an approved customer. These labs will be under the supervision of the Director of the Regional Field Office and will be set up in NASS space. A lab workstation will be set up in a secluded area of the office so that NASS staff conducting sensitive and confidential business cannot be overheard.

Non-NASS Sites - In special cases, NASS may authorize a Data Lab at a non-NASS location. NASS is obligated to keep all datasets under its control, so these labs operate under very strict rules and heavy scrutiny. Remote labs are supervised on-site by approved NASS representative who are trained by NASS and expected to perform their duties as if they were NASS employees.

Access and Confidentiality Certification

Access to the Data Lab and NASS datasets require the approval of the NASS Chairperson, Agricultural Statistics Board. Access is requested using NASS form ADM-042, *Request to Access Unpublished Data*. You will be asked to describe the data source, the data items needed, the level at which you plan to summarize (state, county), location where the data is to be used, and the intended use of the data. Access is evaluated on a project by project basis, therefore, an ADM-042 is required for each project with details on each dataset requested. The request can be sent directly to the Chairperson, Agricultural Statistics Board, to any NASS Regional Field Office Director, or to the Head of the Census Data Section (Datalab@nass.usda.gov).

All data users approved for the Data Lab will be required to have their signature witnessed on a NASS form ADM-043, *Certification and Restrictions on Use of Unpublished Data*. By signing this, you attest that you understand the laws governing confidentiality of NASS data and understand the penalties of violating these laws. The back of the form ADM-043 contains pertinent excerpts of these laws. One ADM-043 is sufficient for all datasets to which access is approved.

Primary Contact

You will have a primary contact person who will be familiar with the NASS data sets and the operation of the Data Lab. For labs at a non-NASS site, the contact is the approved NASS representative. The contact's role is to help you get started with your research by making the dataset available to you, overseeing administrative matters, and arranging any needed disclosure

review.

Data Support

NASS will make available to you the subsets of NASS data that you need once access has been approved. Your lab contact will discuss your data needs in detail to accurately determine the specifications of your dataset. We want to be sure you get all variables needed for your research, but only those variables. No personal identifiers, such as name, address, latitude or longitude will be in the dataset. The default format for these datasets is SAS, a prominent statistical software package. If another accommodation is needed, ask your lab contact.

Allow several weeks from the time you submit your request to when you can expect to get access to the data. This will provide NASS adequate time to review your request and prepare the data set for your use. Also, some lead time may be required to set up a workstation in a lab at a NASS Regional Field Office.

Your lab contact can answer questions about the origins of the data, the processing of the data, variable naming, and other related topics. Your lab contact can also help with questions about the computing environment and dataset structures.

If you wish to bring in data from a non-NASS source, consult with your lab contact. External files must be virus scanned before loading to a NASS computer. Your contact will load acceptable files.

Disclosure Review

As described in the previous section, NASS must perform a disclosure review of all tabulations and statistical measures produced from the data before you remove such material from the NASS facility. This requirement is in accordance with NASS's obligation to protect the confidentiality of data about individual establishments.

NASS employs disclosure limitation methodology commonly used by many Federal statistical agencies. Disclosure methods use two criteria to determine primary suppressions. The first 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 each other's production. The second is a dominance rule which triggers a suppression when releasing a total could allow a very large (and dominant) producer's value to be derived or approximated.

NASS uses the same threshold rule for all programs. We require three operations before a value can be released. That means anytime there are only one or two operations, the value is categorically suppressed. NASS uses different dominance rules in different circumstances. The (n,k) rule invokes a suppression when the top n producers account for k percent of the estimated

total. In other words, a (2,80) rule will suppress a total when the top 2 producers represent 80 percent or more of that total. The p-percent rule requires sufficient protection so that no one could approximate the largest producer's value to within p percent. A 20-percent rule will suppress a total if revealing that total allows someone to estimate the top producer's value to within plus or minus 20 percent. Federal statistical agencies do not publicly disclose the actual values used, as revealing them reduces the protection.

In many instances, a primary suppression requires a complementary suppression for protection. For example, a suppressed value that is one element of a total can be deduced by simple subtraction. Selecting complementary suppressions is much more difficult and time consuming. In some instances, two primary suppressions can serve as each other's complements. However, in most cases, a value that passes the disclosure tests must be sacrificed to protect a primary suppression.

You must observe the following procedures in submitting requests for disclosure review when generating tabulations or model results you wish to remove from NASS. The procedures apply to users of any microdata used in the lab. The following disclosure procedures are designed both to expedite and simplify the process. Before programming for summary statistics or aggregates, review the following information about disclosure. If you have any questions let your lab contact know to avoid additional programming for yourself and NASS.

- (a) To assist with the disclosure review of summary statistics and aggregates, draft the output that you are looking to derive. If you are anticipating 1,000 or more cells of data to be produced let your lab contact know. In these cases it is more efficient for you and for NASS to have NASS summarize the data and utilize an automated disclosure check program.
- (b) Submit all requests for disclosure review to your lab contact who will arrange for the review. The lab contact will forward it to the Washington D.C. Data Lab. The lab contact will provide an estimated turnaround time for the disclosure review. The review time depends on the complexity of your output. We recommend you allow at least 30 days for this process. You should allow time for the disclosure review in your planning - for example, when submitting a paper to a conference.
- (c) Provide an electronic copy of the tables. This allows us to define the data relationships within and across tables to ensure the path of suppressions reaches closure.
- (d) If the number of summary statistics and aggregates is less than 1,000 cells provide a machine readable dataset with the following variables: level of the tabulation (U.S., state, county), the item summarized (description), the value in the table, the number of farms represented by the value, and the reported value of the largest and second largest producers. These data are sufficient to allow a utility program to quickly identify primary suppressions which will greatly reduce the time required for the review. SAS datasets are preferred, but others can be accepted. Consult with your lab contact for advice and assistance. Failure to provide this dataset will extend the disclosure review time.

Note: Farm counts are not considered sensitive data.

- (e) When the tabulation of one variable requires the use of intervals based on the same variable, or another variable, the interval width of the bin must be sufficiently wide. For instance, the midpoint $\pm 25\%$ of the midpoint.
- (f) If we find disclosures, we will instruct you to suppress a portion of your output. You have two options when making suppressions. You may suppress individual table cells by explicitly replacing the cell value with an indicator identifying a suppression. NASS uses a “D” (withheld to avoid DISCLOSURE of an individual operation) in place of the actual number in the affected cells. The other course of action is to “collapse” related cells and publish a group total that passes the disclosure tests.
- (g) For model output, such as linear or nonlinear regression models, disclosure is usually not a problem. However, models with dummy variables that take on a value for a small number of observations may generate a disclosure risk. This is why we check for disclosure failures for all of the regression coefficient “cells”. If we do find a problem, we will require you not to report the values for at least one coefficient. You could still indicate whether they were significant, or perhaps you can report that the coefficient is in a certain range (for example, positive and significant).
- (h) Checks will be conducted by NASS to ensure that your output will not inadvertently disclose a cell previously suppressed in a NASS publication.
- (i) If you do not understand decisions on disclosure, discuss them with the lab contact to try to resolve your concerns.

III. Computing Environment

The computing environment at the NASS Data Lab has been designed to be as user-friendly as possible. We understand that most of the users want to perform their research as quickly as possible, therefore, they do not want to acquire a myriad of new computing skills. Our goal is to provide the most accommodating computer environment within the established computing standards. We want you to be able to manipulate our data as easily and as quickly as possible.

In general, we have designed your computing environment so that data subsets can be easily accessed at your workstation; work can be performed in a graphical user interface environment; and appropriate application software is provided. The objective is to get the data to you quickly and provide the computing tools you need to be successful. We realize that the computing environment is a critical element to the success of your research or analysis.

Workstation Environment

The lab consists of three basic components client hardware, operating system, and application software.

Client Hardware

Lab workstations are comprised of personal computers, which include a central processing unit with data storage, a monitor, a keyboard, and a mouse. We refer to this as the “client”. We try to provide sufficient disk space as each user requires. However, disk space is not unlimited and usage may require some management on your part. In general, if your research is confined by available hardware resources, we will make every effort to accommodate your requirements. You are not allowed to remove any data or output from the client or connect a flash drive or other portable media. Remember, you are never to remove any data or tabulations from the Data Lab without the approval of your lab contact after the disclosure review process has been completed.

For labs in a non-NASS location, the computing configuration has been jointly designed by NASS and the sponsoring agency and may be different from labs in NASS locations.

Operating System

Workstations are set up with the Windows operating system.

Application Software

Labs at a NASS site support SAS, R, and Microsoft Office suite (Excel, Word, and Powerpoint). If you know the specific R packages that you are interested using, let us know before you arrive so we can ensure they are loaded to the client machine. There may be other statistical/econometric packages you may want to use. If NASS already owns valid licenses for these packages, we will try to make the package available to you. If NASS does not support the needed software, you can bring a licensed copy of the software to your lab contact and we will arrange to have it loaded. A licensed copy of any application software used the Data Lab should be left with your lab contact. Pirated software is not allowed on Federal computers.

Operation of the Lab

The computer(s) in the lab may or may not be pre-assigned to a particular user or Agency. Your lab contact will inform you of the machine(s) you can use. There are some basic computing policies which must be adhered to throughout your stay at the lab. These are:

- (1) Do not turn off (power-off) any computing equipment. These devices are powered-on and off by the lab staff.
- (2) Do not change your software configuration without first discussing the rationale for such a change with your lab contact who will work with the computer support staff to oversee any changes.
- (3) Do not directly load software onto a personal computer under any circumstances. Your lab contact will arrange for loading of any licensed software that you require. The Data

Lab must adhere to NASS and USDA policy which states that only NASS approved software can be loaded onto NASS computers.

If you should require additional software, or any other computer related supplies, please speak to your lab contact. Every effort will be made to provide you with the computer resources you need.

- (4) Immediately report any hardware or software malfunctions regardless of how minor.
- (5) Do not connect or portable media to the client to load or copy any files off the client.
- (6) Do not leave the personal computer unattended without securing the system from other potential users.

Computing Support

When you arrive at the lab, your computing environment will be preconfigured by NASS personnel or the approved NASS representative. You will have been added as an authorized user, your password created, your required software will have been loaded, and the data you require will be loaded into a directory for you.

IV. RFO Role and Responsibilities

The Regional Field Office Data Labs (RFOs) are operated on a part time basis within each of the 12 RFOs. These Data Labs allow for shorter commuting distances for many researchers to access data for approved projects.

The RFO is responsible to ensure that all procedures and practices detailed in this handbook are followed at their location.

At each office a stand-alone computer with the software listed in Application Software is available. For assistance with preparing machines contact the NASS Service Desk.

The computer is located within the office where the researcher can easily access their RFO coordinator and is sufficiently removed from the ongoing production activities to ensure the researcher does not have access to additional unpublished estimates or operator information.

Visits to the RFO are scheduled by the RFO staff with the researcher. During the researchers' first visit to the Data Lab for a project, RFO management will provide a confidentiality briefing and witness the completion of NASS ADM-043 form.

The RFO will provide the HQ Data Lab manager notification when output is ready for disclosure review and make the files available for HQ access.