NATIONAL
AGRICULTURAL
STATISTICS
SERVICE

U.S. Department of Agriculture
1400 Independence Avenue, S.W.
Washington, D.C. 20250-2000

SAP Research Application Guidelines

Persons wishing to conduct research on restricted National Agricultural Statistics Service (NASS) data in a virtual data enclave must submit a research proposal to NASS through the Standard Application Process (SAP). The purpose of this document is to provide guidance for the various sections of the SAP. Generally, this document applies to datasets that list NASS as a Source in the SAP metadata catalog.

NASS and designated agents are required to protect the confidentiality of data collected under the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2018, Title III of Pub. L. No. 115-435, codified in 44 U.S.C. Ch. 35 and other applicable Federal laws. All research must have a statistical purpose and may not be used for regulatory, enforcement, or investigative purposes.

The Five Safes framework breaks down the decisions surrounding data access and use into five related but separate dimensions: safe projects, safe people, safe data, safe settings and safe outputs.

More information about accessing NASS restricted microdata is available on the NASS website.

Table of Contents

SAP Research Application Submission Requirements 3
Data 3
Research Team 3
Research Description 4
Project Title 4
Project Duration 4
Funding 4
Timeline 4
Research Question(s) 5
Demonstrated Need 5
Study Population 5
Project Abstract 5
Variables Requested 5
Time, Geographic, and Other Units Requested 6
Provisioned by selection requirements 6
Zip Code & Latitude/Longitude 6
Work Location(s) 6
Data Linkages 6
Operation Linking Across Time using Census of Agriculture 7
Operation Linking Across Census of Agriculture and Follow-on Surveys 7
Operation Linking Across Census of Agriculture and ARMS 7
Linking to User Provided Data 7
User-Provided Data 7
Software Requirements 8
Methodology 8
County level analysis 9
Watershed (HUC6) analysis 9
Producer Characteristics & Demographic Analysis 9
List of References 9
Project Products 10
Requested Output 10
Agency Benefits 11
Documentation 11
Multiple Institutions 11

SAP Research Application Submission Requirements

Researchers who wish to conduct research using NASS data are required to read, understand, and comply with the NASS Data Lab Handbook. The Handbook contains valuable information on the secure data enclave, disclosure methodologies utilized by NASS, and the requirements all researchers must adhere to when conducting their research and requesting data for export.

One member of the research team is responsible for submitting the proposal online using the Standard Application Process (SAP). The submitted research proposal will be reviewed by NASS. Applications for data assets that have multiple authorizing agencies, such as ARMS, will be reviewed by all authorizing agencies. The SAP will ask for information about the research team and the proposed project. Question verbiage and layout in the SAP may not be exactly as shown below. If researchers are uncertain about how to respond to an item, refer to the RDG User Guide, this guide, or contact the NASS Data Lab and Data Access Group at SM.NASS.Data.Lab@usda.gov.

Data

The first step to beginning a project proposal using NASS data is to select a dataset by clicking the "Request Access" button in the SAP metadata catalog. Researchers are encouraged to review all information available for each dataset in the metadata catalog on the SAP website and reference it when selecting the data and years available for each dataset. Multiple datasets can be included in the application. To check whether datasets in the SAP metadata catalog can be accessed together, click on "Data Access" and verify that the Access Modality indicates USDA Virtual Data Enclave. After selecting all datasets needed for the project, click "Start Application."

The SAP Metadata is updated as data products or survey years are ready for Data Lab use. Data collected by NASS that are not listed in the SAP metadata are unavailable for researcher use.

Research Team

The proposal must include all members of the research team, including collaborators who do not plan to access NASS data directly via an enclave account, but will participate in planning, oversight, or dialogue about the project prior to disclosure review. List the name of the principal investigator (PI) and all other researchers associated with the project. NASS does not allow students to be designated as the PI. Students seeking data access must designate their primary advisor as the PI. Refer to the NASS Data Lab Handbook for eligibility criteria. Foreign national researchers are generally permitted to access NASS data, but under the National Farm Security Action Plan released on July 8, 2025, foreign nationals from *Countries of Particular Concern* will not receive security clearance.

All researchers, including U.S. citizens, must access NASS data from within the United States.

For each member of the research team, the application form requires full legal name, institutional affiliation, title (faculty, grad student, research scientist, etc.), institutional email address (not personal),

phone number, citizenship status, and Special Sworn Status. All researchers accessing NASS confidential data, through a direct account or through project collaboration, must answer Data Access = Yes and complete security protocols. For student team members, a graduation date must be entered.

If new collaborators join or researchers leave the project after the proposal has been approved, the NASS Data Lab and Data Access Group will assist with adding or removing them from the project via amendment.

Research Description

This portion of the project application will focus on the project details. Each field or section must be written so that it can be understood by a competent statistician who is not necessarily a specialist in or on the topic within the field of study. Keep in mind that the audience for a NASS research project proposal is not the same audience researchers typically address when writing research proposals or journal articles. NASS has its own set of expectations and requirements, and therefore researchers often find that they must make significant revisions to preexisting proposals written for other purposes (e.g., grant applications).

Project Title

Provide a descriptive title for the project.

Project Duration

NASS encourages researchers to carefully assess the time period for which they request data access and ensure the intended research can be completed in this timeframe. The maximum project duration is 36 months. Requests for extensions beyond the specified end date undergo careful scrutiny and generally are granted only for circumstances beyond the control of the researchers (e.g., pandemic). Note that researchers may discover that it takes longer than anticipated to become familiar with the data and computing environment.

Funding

List sources of funding for data access fees, as well as funding for the research project overall. NASS does not charge researchers for use of the data but does require the researcher to pay for services of the data enclave provider. The enclave provides a secure environment for researchers to access data while administering confidentiality protections. General fee information for enclave access can be found on the NASS website.

Timeline

Upload a table with a timeline, including a column for each year and a row for each task. A generic timeline template is available in the SAP application for download. The timeline must include activities such as cleaning data, merging data, and conducting specific analyses, as well as an estimated timeframe for

outputs, such as disclosure review, publications, presentations, or any public use products. Every timeline must incorporate a task for disclosure review. Refer to the timeline below as an example.

Task	**year**		
	1	2	3
Compile, clean, and merge data, review outliers	X		
Initial analysis: hypothesis test and summary statistics	X		
Intermediate analysis: code and run model		X	
Final analysis: robustness checks and alternative data		X	
Draft paper of research findings		X	
Send export request to NASS for disclosure review		X	
Submit to journal		X	
Respond to peer review			X
Publish journal article			X

Research Question(s)

Provide an overall description of the research question(s). This section must describe the purpose of the project including any key concepts that will inform the research methods or models and a general description of the sample(s).

Demonstrated Need

Discuss why the project needs non-public data. Explain in detail why the project's research questions can only be addressed using the requested restricted-use microdata. Be specific.

Study Population

Briefly describe the study population or universe and how it relates to the research question.

Project Abstract

This section must be 150-250 words in length, reference major datasets and key methods; discuss expected findings. Researchers may include one or two sentences about why the research is important to their field(s) of study. It must capture the essence of the project proposal, similar to a scientific journal abstract.

Variables Requested

NASS requires that a variable selection spreadsheet be submitted for applications requesting access to Census of Agriculture data for the years 2007 and later. It is not required for any other datasets. The spreadsheet of available variables can be found on the NASS website or downloaded directly through this link. Follow the directions on the 'Information Sheet' tab to select project variables. Variables requested need to be clearly justified in the project proposal. Write-in variables cannot be added to the spreadsheet and will not be approved.

Since the SAP application does not currently allow spreadsheets to be uploaded, this spreadsheet must be emailed to SM.NASS.Data.Lab@usda.gov after you submit your application. Include the application number in the subject of the email. Delay in receiving the spreadsheet will result in the application being returned to the researcher for revision and resubmission, delaying review of the application.

Time, Geographic, and Other Units Requested

This field is required for all NASS applications. Generally, for each requested dataset, select the provisions of data needed within the data options for each data asset. The selected provisions must be clearly justified in the methodology section.

Provisioned by selection requirements

Most data assets, like the Census of Agriculture and ARMS, require additional selections to be made related to provisioning by time/geography or time/commodity. These selections are required for application approval.

Zip Code & Latitude/Longitude

Requests for zip code or latitude/longitude coordinates will not be approved for any research projects.

Work Location(s)

Only "USDA Virtual Data Enclave" can be selected for projects that access NASS data.

Data Linkages

Upload a document summarizing how all the data fit together, how they will be linked, and their expected overlap, including expected sample sizes (if applicable). State the datasets to be linked, the unit at which the linkages will occur (e.g., operation, geographic, etc.), the purpose of the linkage (e.g., geographic context), and information as to how the linkages are to be performed.

It is important to describe how and at what unit of observation or aggregation datasets will be linked. Linking survey data to other survey data may lead to small sample sizes (and therefore potential disclosure issues) if the survey samples do not significantly overlap. Also, linking a single survey over time can lead to similar concerns. When linking confidential data to externally provided contextual variables, detail the level of geography (e.g., county, state, etc.) at which this link will be performed.

An application must describe all planned linkages including operation linking within NASS data assets as well as any information related to how NASS data assets will be linked to user provided data.

Operation Linking Across Time using Census of Agriculture

Creating a panel dataset by linking operational responses over time will only be approved for Census of Agriculture requests. The Data Linkages section of an SAP application must indicate intention to link operations from the same survey across time. Panel datasets created from the Census of Agriculture will only be able to link data from in-scope responding operations. The panel dataset cannot be used to accurately measure operation entry or exit from the agricultural sector due to a researcher's inability to distinguish between survey non-response and/or presence on the sample frame.

Linking at an operation level is possible only with data from the 2007 Census of Agriculture onwards due to changes to the NASS List Frame maintenance procedures. Provisions selected in the Time, Geographic, and Other Units Requested section of the Research Description must reflect applicable years only.

Operation Linking Across Census of Agriculture and Follow-on Surveys

It is possible to link the Census of Agriculture positive responses to positive responses from other Follow-On surveys like Census of Aquaculture and Irrigation and Water Management Survey within the same 5-year census cycle. The resulting linkages will be affected by survey non-response or changes to the operation between the two surveys. Due to low match rates, fitness of use is reduced and therefore linking is discouraged.

Operation Linking Across Census of Agriculture and ARMS

There is no benefit to linking Census of Agriculture and ARMS Phase 3 data due to NASS' data collection strategy. During a Census year any respondent to ARMS Phase 3 is not required to fill out a Census questionnaire to reduce respondent burden. Instead, their ARMS Phase 3 questionnaire responses are copied directly into a Census response.

Linking to User Provided Data

All NASS data assets do not contain any Personally Identifiable Information (PII) such as an operation's physical address or latitude/longitude coordinates. User provided data is usually aggregated and linked by the researcher using the lowest geographical indicator provided in each NASS data asset.

User-Provided Data

If the project team plans to utilize non-NASS data in the research project the application must include a comprehensive list of outside datasets. Enter the dataset name, a short description of the data, and the approximate size of the file. The application will ask for the data source, whether the data is publicly available or proprietary, and if it is anonymized. If available, include the link where the data can be downloaded.

Indicate if user-provided data will be linked during the analysis. The Data Linkages section of the application must detail how this data will be connected to the requested NASS data assets.

If the application is approved, the research team will be required to provide the data to NASS prior to workspace access. Future requests for user-provided data not listed in the proposal will require additional justification and approval and will result in an ingestion fee to the data enclave provider.

Software Requirements

The virtual data enclave has statistical software ready to use in the computing environment, such as R and Python. Available software applications can be found in the SAP metadata catalog; click on Data Access, then on USDA Virtual Data Enclave under Access Modality. Consult with the NASS Data Lab and Data Access Group at SM.NASS.Data.Lab@usda.gov about other software needs. If the project team requires statistical software other than that which is currently available in the USDA Virtual Data Enclave, enter details in this section. Accommodation for unique software requirements may be available; provide justification, priority, and the inability of available packages to do the work. Fees will apply. The virtual data enclave environment does not have internet access which may affect the functionality of any unique software requirements.

Methodology

This section describes the statistical equations that will be estimated, what key variables are needed and how they will be measured, how the data will be used, and how all datasets fit together. It is important in this section to describe the methods used to answer the research questions.

All requested datasets, including user-provided datasets, need to be discussed in the methodology section. The purpose for each requested dataset and its role in the analysis must be clear. Data crafting is an essential element of research design. The proposal needs to be as detailed as possible in describing how measures will be constructed and how the data will fit together to construct the focal sample(s). Generally, this is accomplished by discussing how variables are constructed and from which datasets they are retrieved.

It is important for researchers to convey to the application reviewers that they have read the publicly available documentation about the datasets they are requesting and have thought through implications of the data in relation to their research approach. Sources of information about datasets that are linked to the SAP metadata catalog include the website link for general survey information and the supplemental documentation, including survey questionnaires and data dictionaries, which can be downloaded from the catalog.

The unit(s) of observation for the analysis and the groups of units for which researchers will carry out analysis are both significant. For each, be specific. Is the unit of observation a farm operation? Alternatively, will the unit of observation be at a more aggregated level (e.g., state)? Will the analysis use some combination of units? Additionally, for what groups of units (what "levels") will researchers specify

their models? Will researchers estimate models at national or sub-national levels? If so, at what levels? In addition, will researchers run separate models by demographics, crops, etc.? If the project uses user-provided data, how big are the samples that will be analyzed? Information that describes the units of analysis and the groups of units/levels of samples the project will be using are necessary for NASS reviewers to assess both the feasibility of the project and the risk of disclosing confidential information.

Write out the equations that will be estimated. Researchers do not need to give the exact functional form they plan to use, but they must write out, in as much detail as possible, the general equation and relevant variables and discuss how the data will fit into the equation. Researchers do not need to list every specification, nor every variable they will be using from each dataset, but they must discuss how the left- and right-hand sides of each equation will be measured (i.e., if they are dichotomous, categorical, or continuous, etc.). Specify the dataset(s) from which these measures will come and how they will be constructed.

County level analysis

County level analysis may only be approved on research projects using the Census of Agriculture. All other NASS data assets are sampled and weighted for state or region level aggregation. A geographic indicator for county may be included in the microdata but will only be permitted for linking purposes. Apart from the Census of Agriculture, any research plans for analysis or output lower than state level will be rejected due to concerns over fitness of use and disclosure.

Watershed (HUC6) analysis

Six-digit Hydrologic Unit Code (HUC6) indicators are only available in Census of Agriculture and Irrigation and Water Management Surveys. Any intended use of HUC6 must be documented thoroughly throughout the application with particular emphasis on the methodology and output section details. Requests for HUC6 analysis undergo additional scrutiny due to fitness of use and disclosure concerns. For Census of Agriculture, the HUC6 variable must be selected in the variable request spreadsheet.

Producer Characteristics & Demographic Analysis

All persons reported on the Census of Agriculture questionnaire must be included in any analysis plans that involve producer characteristics and demographic analyses. Applications that propose methodologies focusing solely on primary operators or principal producers will not be approved.

List of References

Provide a brief discussion of the relevant literature. Provide a targeted discussion of the literature that focuses on the key datasets as well as the key concepts to be examined in the analysis. This will help demonstrate scientific merit of the research question(s) as well as demonstrate a deeper knowledge of the requested data. Include a list of publications referenced in the proposal.

Project Products

Describe the deliverables that the research team anticipates producing, such as journal articles, conference presentations, technical reports, PhD dissertations, government reports, and other expected products. Include the names of journals targeted for publication.

Requested Output

In this section, researchers must describe the output they expect to ask to release from the secure data enclave. This description is important in assessing both the substance of the proposal and the risk of disclosing confidential information. Output from the enclave must focus on model-based results (regression coefficients, standard errors, and the like) and be consistent with the approved project objectives. Data mining is not permitted. Summary statistics (variable means, etc.) are allowed, but only to the extent that they support model-based output. For example, the kinds of output expected to be released are regression estimates and tables similar to those found in an article in a peer-reviewed academic journal. Researchers who desire large volumes of tabular output must request a special tabulation from NASS rather than requesting access to microdata. To protect sensitive and confidential data, all products and research results created in the secure data enclave are considered confidential until they have undergone NASS disclosure review and been approved for release.

NASS looks for several things in evaluating proposals for disclosure risk. The project application must clearly state the proposed methods and output. Reviewers understand that, because the research team is conducting research, it is unlikely that they know all the details in advance. Nevertheless, reviewers need enough detail to assess whether the proposed project can succeed without posing undue risk. Reviewers will examine the output section for details on:

- **Possible "thin data":** Reviewers need to determine whether there may be possible "thin cells" in the output; for this purpose, a cell is the group of observations (e.g., farm operations) underlying any estimate researchers may release. To evaluate this, reviewers need clear and accurate information on the types of output to be requested (e.g., model tabulations, graphs), the units of analysis underlying the output (e.g., all farm operations, or groups of these), and the groups (levels) for which researchers will request output (e.g., crops, geography, demographics, and possible crosses of these). For example, detailed demographics-by-geography cells become thin very rapidly which will raise concerns for fitness of use as well as disclosure.
- **Modeling intentions**: Reviewers need to know whether variables included in models are discrete or continuous; both are allowable, but discrete variables (especially dummy (0, 1) variables) define "cells" that reviewers must carefully consider. Note that including indicator variables in the analysis creates additional "cuts" at the sample (i.e., cross tabulations), but only if the estimated coefficients associated with the indicator variables are reported. Because of this potential issue, the review team encourages including statements in the proposal, as appropriate, for example: "Specifications will include detailed farm type controls. However, we will not release these fixed effect coefficients; we will only note in our results tables that we included them."

- **Plans to produce Tabulations/Graphs/Maps**: Reviewers need to know whether researchers will request tabulations, graphs, or maps. These products can be particularly difficult to check against required disclosure standards.
- **Lowest geographical aggregation**: Reviewers need to know whether researchers plan to produce sub-state estimates. Refer to the Methodology section for more information on sub-state aggregation limitations.
- **Statistical Output**: Reviewers will also look for types of output that are relatively unfamiliar, such as from new statistical techniques. They must try to assess their disclosure risk and welcome any insights from researchers.

Disclosure risk is a complex topic, and researchers will need to discuss their proposed output with their research team as they develop their proposal.

Application approval is not a guarantee that all intended output presented in the application will be approved for export from the secure data enclave.

Agency Benefits

This section is not required by NASS for application approval and N/A can be placed in the response box.

Documentation

NASS requires one document beyond the application itself. The MOU is an agreement between NASS, the institution/agency, and the researchers. The MOU must be signed by the Principal Investigator/Project Lead who is an employee of the institution and by a designated Senior Official of the Organization who is authorized to enter into contractual agreements. Signatures must be digital or wet. No electronic or typed signatures are accepted. Modifications to the MOU provided in the SAP will not be permitted under any circumstances. The project lead on the MOU cannot also sign as the Senior Official.

The MOU will be signed by USDA officials and returned to the Project Lead once the project is reviewed and approved.

Multiple Institutions

If multiple institutions are participating in the project, a separate MOU is required and must be uploaded into the application for each organization or institution. Each MOU will require the signature of a faculty member/employee from the institution participating in the research project as well as a signature of a Senior Official from the institution authorized to bind the institution to the terms laid out in the MOU.