The General Social Survey GSS

NORC at the University of Chicago

America's Research Project Since 1972

Updated: April 1st, 2024

Obtaining GSS Sensitive Data Files

General Social Survey (GSS) Sensitive Data are made available to researchers to use for a limited period under a special Sensitive Data Set License Agreement with NORC. The GSS takes its promise of confidentiality to its respondents very seriously and is the basis for the Agreement process. Under Agreement, the GSS will provide data on detailed geographic information (e.g., state, county, primary sampling unit, and census track) or mortality-related information (e.g., year of death, cause of death), but in no circumstances will individually identifying information (name, address, etc.) be provided.

An overview of the application process for obtaining the GSS Sensitive Data is outlined below:

1) You submit an application for GSS Sensitive Data, following the guidelines provided in the GSS Sensitive Data Application section below, to NORC by email to GSS@NORC.org. Your application will include:
    a) **Research Plan:** The Research Plan must fully specify all the research that is to take place using the data and must be project specific. Research Plan must also include:
        i. **Human Subjects Review Clearance:** Approval from your institution's human subjects review board (e.g., IRB) for the research described in the Research Plan.
        ii. **Curriculum Vitae** for each participating research staff.
    b) **Sensitive Data File Request Form:** This form specifies which datasets and variables you are requesting.
    c) **Sensitive Data Protection Plan:** The Sensitive Data Protection Plan details how you and your institution intend to keep GSS Sensitive Data safe following the guidelines provided.

    Your application will then be assigned to one of our sensitive data reviewers.
2) Once your application is approved, we will send you the Sensitive Data Set License Agreement via email. You will return the signed Agreement along with a license fee of $1500 USD.
3) We will send you a fully executed Agreement, a Secure File Transfer Protocol (SFTP) location with the requested data, and all necessary documentation. You can then merge the GSS Sensitive Data with your own GSS public data (and other datasets as described in your approved application).

The process to obtain GSS Sensitive Data can take several months. To avoid delays, we ask potential users to carefully review the details on the application process contained with this document. To request further instructions, any of the documents mentioned above, or questions regarding the process in general, please contact us at GSS@NORC.org.

Effective April 1, 2024, NORC has increased the data license fee for GSS Sensitive Data from $750 to $1,500 for a 3-year data license. NORC is committed to making most of the GSS data publicly available at no cost to researchers. However, there are some variables that NORC withholds from the public use files as they contain information that may lead to someone in the GSS being identified (for example, detailed geographic information). NORC has increased the application fee for the GSS sensitive data license to $1,500 for a 3-year data license. This will help NORC's GSS team process applications, make improvements to facilitate the data exchange, as well as better safeguard the data.

Our new updated data license fees will be used to safeguard GSS respondent confidentiality by reviewing IT protocol for applicants, review data output from applicants, and protocols for ensuring applicants follow the terms of the license. As part of this new license process, we are allowing more flexibility to meet IT security requirements with the goal all applicants will meet the GSS IT security objectives.

We provided users with notice of this fee increase as of January 26, 2024. If this change impacts an upcoming sensitive data application, please reach out to the GSS team at GSS@NORC.org and we will do our best to work with you.

GSS Sensitive Data Products Available

GSS Geographic Data: For cross-sectional data, these include state (1973-2022), state at age 16 (1978-2022), primary sampling unit (1973-1993), county (1993-2022), and census tract (1998-2022). State, state at age 16, county, and census tract are also available for the GSS Panel sample whose original sample years were 2006, 2008, and 2010. Please note that 1998 census tract was assigned according to 2000 US Census.

National Death Index (NDI): For cross-sectional data, these include year of death or age at death (1978-2010) and cause of death (1978-2010) as of 2014. Vital status as of 2014 is available in the GSS public data (variable name: DEATH).

GSS Sensitive Data Application

The GSS Sensitive Data files are made available to researchers at an institution under special agreement with NORC. The GSS, like most interview-based surveys, has promised confidentiality to its respondents. This promise is taken very seriously and is the basis for the Agreement process. No individually identifying information (name, address, etc.) will be provided, even under Agreement. To ensure potential users intend to honor this commitment in their research, we require potential users to apply for GSS Sensitive Data describing their research plan and how the investigators intend to protect the data.

The required applications materials for requesting GSS Sensitive Data are outlined below:

1) **Research Plan:** The Research Plan must fully specify all of the research that is to take place using the data and must be project specific. The plan must specify which years you intend to use and justify the inclusion of variables you are requesting. The plan must include a list all participating research staff who will work directly with the GSS Sensitive Data. It is not permitted, for example, for a faculty member to obtain the data for his/her own research project and then "lend" the data to a graduate student to do related dissertation research, unless this use is specifically stated in the research plan. If you plan to link GSS Sensitive Data with other data, you must include specifics about the other data set(s) and at what levels the data will be linked. Please include a Research Summary, a one paragraph description of the project and how you will use GSS Sensitive Data. The Research Plan must also include:
    a) **Human Subjects Review Clearance:** Obtain Human Subjects Review Clearance from the appropriate body at your institution (e.g., Institutional Review Board, or IRB) to conduct the research described in your Research Plan. This approval or waiver should be appended to the Research Plan.
    b) **Curriculum Vitae:** Appended to the Research Plan should be a copy of the curriculum vitae or academic resume for all participating research staff.
2) **Sensitive Data File Request Form:** The request form is available at https://gss.norc.org/Documents/other/Sensitive_Data_File_Request_Form.pdf. You will need to specify the data format and the specific datasets and variables you are requesting. These selections must be consistent with your Research Plan. Data are available in SAS, SPSS, and Stata formats.
3) **Sensitive Data Protection Plan:** Study the Criteria for Sensitive Data Protection Plans section below and investigate mechanisms that are available to you to meet its requirements at your site. Please consult with your institution's IT security team to ensure reasonable data security measures can be taken. The plan must meet these requirements, or your reviewer will require you to resubmit the plan. If the computing systems/environment change, a new Sensitive Data Protection Plan based on the new system must be submitted and approved.

All of the requested application materials should be emailed to GSS@NORC.org.

Multi-institutional research teams: We understand that there is a lot of collaboration between research institutions. However, a critical component of the GSS Sensitive Data process is the Sensitive Data Set License Agreement. **NORC will only enter into an Agreement with a single institution.** This means that if your project has data users at multiple institutions, your team will need to have multiple Agreements in place – *each with its own license fee* – which will require their own distinct Sensitive Data Protection Plan and institutional approval, and independent data deliveries. This may increase the time to obtain data as well as project budgets. Please carefully consider before applying whether you need data users at multiple institutions.

Application Process

Institutional bureaucracies being what they are, the process to request and receive GSS Sensitive Data can take several months. In order to avoid needless delays, we recommend that potential users of these Sensitive Data files take the following steps, in the following order, submitting all documentation at one time:

1) Application Review: After receiving your application, we will assign one of our sensitive data reviewers to review your application. Depending on GSS priorities and any backlog of requests, this process may take up to 4-8 weeks. Please be patient. During this review, your assigned reviewer may ask for clarifying information, updated documentation, or provide guidance on how to have your application meet required expectations.
2) GSS Sensitive Data Set License Agreement: Upon approval of the application materials, a GSS Sensitive Data Set License Agreement will be sent. Please note the following:
    a) A Sensitive Data Set License Agreement is between NORC and one institution. If your project has data users at more than one institution, you will need one Agreement per institution. Please carefully consider before applying whether you need data users at more than one institution.
    b) Agreements must specify the name and number of all data users at the receiving institution. All data users and an authorized administrator from the receiving institution must sign the Agreement.
    c) There must be a Co-Investigator in situations in which the Investigator does not have a full-time permanent faculty-level appointment at the institution where the research will take place (e.g., where the Investigator is a visiting scholar or a graduate student). The Co-Investigator must be a PhD level, full-time faculty member at the receiving institution.
    d) There is a requirement that the Investigator(s), the Co-Investigator and the receiving institution assume liability for any violations of the agreement by any person at the receiving institution. If the institutional representative has issues with the liability language in the agreement, please contact the GSS representative immediately. THE AGREEMENT WILL NOT BE APPROVED WITHOUT A LIABILITY SECTION.
    e) Researchers agree to be good stewards of the data and follow relevant cell suppression and disclosure review procedures. For more details, see the section on Data Stewardship and Publication.
    f) All original data files must be returned to GSS or destroyed within the specified time limit (3 years, with an option of an extension). All files and paper printouts containing GSS Sensitive Data or data derived from the GSS Sensitive Data must be destroyed or returned prior to the completion of the agreement. A Certificate of Compliance, stating that all GSS Sensitive Data have been returned or destroyed, must be signed and returned before the agreement is closed. Extensions to the timeframe stated in the agreement will be addressed on an as needed basis.
    g) Applicants will need to combine the signed Agreement with the requested attachments including your Research Plan with IRB approval and CVs, Sensitive Data File Request Form, and Sensitive Data Protection Plan.
    h) If the Investigator changes institutions, the current Agreement is no longer valid and a new Agreement must be completed. Also, the original data and any analysis files must be returned to the GSS until the new Agreement is established. If the Investigator wishes to continue to use the data, a new Sensitive Data Protection Plan and Agreement are required. A new license fee must be submitted with the new Agreement before the original data will be returned.
3) License Fee: Along with the signed Agreement, a $1500 USD non-refundable license fee covers the expense of creating, delivering the data files and documentation, for up to four hours of consultation with the GSS staff, data disclosure review for results intended for publication, and the cost of administering the Agreement. This fee can be sent via check, ACH payment, or wire transfer. Please note international transfer fees may apply.
4) Data Delivery: Once the license fee has been successfully deposited, we will provide the Investigator unique login information to obtain the GSS Sensitive Data through a Secure File Transfer Protocol (SFTP). Your institution will need to ensure an SFTP application, like WinSCP or FileZilla, is available to access the SFTP folder with your requested data as an encrypted, password-protected zip file.

Data Stewardship and Publication

The Research Plan and Sensitive Data Protection Plan guide how researchers analyze GSS Sensitive Data. Please note that the GSS data are not designed to be representative of smaller geographies (e.g., states, metropolitan statistical areas,

counties). At the completion of planned analyses, researchers should be mindful of the publication guidelines for GSS Sensitive Data:

1) Cell Suppression Policy: Any documents (manuscript, table, chart, study, report, etc.) created using the GSS Sensitive Data Set must adhere to the following disclosure procedures:
    a. Unweighted cell counts <= 5 must be suppressed.
    b. If a suppressed cell is the only cell in a row or column that is suppressed, then at least one additional cell must be suppressed, or the value of the sensitive cell should be calculated by subtraction from the row or column total.
    c. Categories that are a composite of other categories (e.g., "Other") are exempt from this suppression.
    d. Suppression can also be achieved by combining categories that do not meet the suppression criteria with other categories until the threshold of >5 (unweighted) is met.
    e. Any row or column percent equal to 0% or 100% must be suppressed.
    f. These procedures apply to any table or graph derived from a table.
2) Geographic Identifiers: Researchers are not standardly permitted to publish any state, county, primary sampling unit, or census tract names or identifier without prior express written permission, because these identifiers are considered "sensitive data" (see Definitions). NORC may grant permission to publish geographic names following a data disclosure review, including the reason for the disclosure and tables proposed for publication.
3) Data Disclosure Review: This review is not intended as a peer-review process, but to ensure that NORC, as stewards of the data, can ensure that only aggregated results are published, the cell suppression policy is followed, and geographic identifiers are not included, unless allowed per the Agreement.

Criteria for Sensitive Data Protection Plans

The Sensitive Data Set License Agreement requires that potential investigators submit a Sensitive Data Protection Plan for approval by the GSS staff. This requirement is part of the effort to ensure that the promise of confidentiality to the GSS respondents is kept and that no persons other than those authorized by the agreement (the named Investigator(s), Co-Investigator, and Research Staff) have access to the contents of the GSS Sensitive Data files. Applicants should consult with their institution's IT security team to ensure reasonable data security measures can be taken.

Definitions

In drafting the Sensitive Data Protection Plan, keep in mind the following definitions:

1) "Sensitive Data" includes any data from the GSS that might compromise the confidentiality of respondents to those studies. Specifically, it includes any data file that, for either individuals or families, includes:
    a) Identification numbers or demographic information (such as month and year of birth, age, ethnicity, occupation, industry, gender, etc.);
    b) geographic identification of areas smaller than census Division, including, but not limited to state, county, minor civil division, primary sampling unit (PSU), segment, city, place, zip code, tract, block numbering area, enumeration district, block group, or block;
    c) any variables or fields derived from the data mentioned in items a)-b) above, including data linked to a GSS dataset using the data mentioned in items a) and b) above as linking or matching variables.
2) "Authorized Person" includes the named Investigator(s), Co-Investigator, and Research Staff. With the partial exception of some computing center personnel noted below, all other persons are referred to as "unauthorized person".

General Requirements

GSS Sensitive Data must be stored using reasonable security practices and procedures to ensure data is only provided to those who require access to complete what is described in the Research Plan ("need-to-know") granting the minimum system resources and authorizations to perform ("least privilege"). This may include, but is not limited to, storing the GSS Sensitive data on a removable storage device kept in a locked storage location (i.e., in a locked desk, in a locked office), on a password-protected hard-drive of a free-standing desktop PC on a *non-networked* directory, or on a restricted network segment.

The Sensitive Data Protection Plan must address each of these issues:

1) A description of the computing environment in which the named research staff will be managing and analyzing the data. For each item of computing equipment and removable storage devices to be used (hard disks, CPU, disks, printers, flash drives, etc.), describe the following:
    a) their location;
    b) persons who have physical access to them;
    c) the security provisions that restrict access to use of data on the system(s), such as locked doors, locks on equipment, passwords, etc.; and
    d) the routine procedures for making backup copies of data files on any storage devices.
2) Include a description of how access to GSS Sensitive Data files will be limited to only authorized personnel. The description should include details of security measures, such as passwords and read/write access to the relevant files.
3) The plan must indicate how routine backups of the Sensitive Data will be prevented. The plan must include a statement that no more than one backup copy will be made of any file containing Sensitive Data and this copy will be destroyed (written over or otherwise made unreadable) prior to the return of the GSS Sensitive Data.
4) A description of how access to any removable storage devices such as flash drives or external hard drives will be restricted. Include such information as where the flash drives/external hard drives are physically located and how physical access to them is to be restricted, including provisions for storage in locked cabinets when not in use. If you will not be using removable storage devices, clearly state this in your plan.
5) Include a description of how access to paper printouts containing Sensitive Data will be restricted. The GSS Sensitive Data Plan reviewers strongly recommend against the creation of any paper printouts containing Sensitive Data. If paper printouts must be used, the plan must clearly state why no other storage media could be used. Additionally, storage issues must be addressed, such as locked storage; how the printouts will be kept from the vision and reach of unauthorized persons when in use; and how the printouts will be destroyed (made unreadable). If printouts will not be used, simply state this in the plan.
6) Treatment of data derived from the sensitive data: We require a clear statement that you will treat all data derived from sensitive data in the same manner as the original sensitive data, and that you understand that data derived from sensitive data includes, but is not limited to:
    a) subsets of cases or variables from the original sensitive data;
    b) numerical or other transformations of one or more variables from the original sensitive data, including sums, means, logarithms, or products of formulas; and
    c) variables linked to another dataset using variables from a GSS sensitive dataset as linkage variables.

    Aggregate statistical summaries of data and analyses such as tables and regression formulae are not “derived variables” in the sense used in this Agreement and are not subject to the requirements of the Sensitive Data Protection Plan and Agreement.
7) Indicate which other GSS and non-GSS datasets, if any, you intend to link to the GSS Sensitive Data you are requesting and include a clear statement that you will not perform linkages to any other datasets. Your statement must include recognition of the following rule that no GSS sensitive dataset may be linked to any other dataset without the explicit written permission of GSS.