



**National Science Foundation
(NSF)**

Scientists and Engineers Statistical Data
System (SESTAT)
Privacy Impact Assessment

Table of Contents

1. Contact Information	1
2. General System Information.....	1
3. Data in the System	2
4. Attributes of the Data (use and accuracy)	3
5. Sharing Practices	4
6. Notice to Individuals to Decline/Consent Use.....	5
7. Access to Data (administrative and technical controls).....	5
8. Privacy Analysis	6

Revisions

Revision Number	Author	Date	Description
Template Update	M Tillotson	July 2008	Updated the PIA template
2.0	A Kifle	April 2010	Updated Template
3.0	A Kifle	March 2011	Update
4.0	M Tillotson; M Powell, S Cohen	September 2011; October 2011 OGC and NCSSES	Update

'
'

Privacy Impact Assessment

1. Contact Information

- a. Project Manager/System Owner(s):
- **R. Keith Wilkinson**, Senior Analyst, Social Behavior and Economic Sciences (SBE), National Center for Science and Engineering Statistics (NCSES), rwilkins@nsf.gov, (703) 292-7801
 - **Steven Proudfoot**, Survey Statistician, Social Behavior and Economic Sciences (SBE), National Center for Science and Engineering Statistics (NCSES), sproudfo@nsf.gov (703) 292-4434
 -

2. General System Information

- a. Name of system or electronic collection of information: '
• Scientist and Engineers Statistical Data System (SESTAT). '
<http://www.nsf.gov/statistics/sestat/sestat.html> '
- b. Description of system or electronic collection of information: '
• The SESTAT Data Tool provides on-line access to three individual survey data collections, collectively known as the SESTAT surveys, along with access to an integrated set of survey data.
- c. What is the purpose of the system or electronic collection of information? '
• SESTAT Data Tool provides a web-based access for the analysis of scientific and engineering labor force data; SESTAT is a source of information for formulating NSF, Federal and state government agencies policy.
• The NCSES website provides for public access to demographic survey data collected by NSF Social Behavior and Economic Sciences (SBE), National Center for Science and Engineering Statistics (NCSES) division.
• NCSES offers a variety of publications that summarize SESTAT surveys on the training and employment of scientists, engineers, and technical experts.
• The SESTAT Data Tool offers a searchable metadata explorer that provides access to a wide variety of information about the surveys and the data.
- d. Requested operational date?
• The SESTAT Data Tool has been operational since 1995. The tool provides access to survey data for 1993, 1995, 1997, 1999, 2003 and 2006.

- A version of the SESTAT Data Tool meeting Rehabilitation Act Section 508 standards is available through *Text Only* links on the bottom of SESTAT web pages.
- e. Does this collection create a new Privacy Act System or is this information collection covered by an existing Privacy Act System? If so, what is the name of the current Privacy Act System?
- NSF 43 - Doctorate Work History File
 - NSF 58 - National Survey of Recent College graduates and Follow-up File.
- f. What specific legal authorities, arrangements, and/or agreements require the collection of this information?
- NSF 43 - Authority for collection of information from scientists and engineers is provided by the National Science Foundation Act of 1950, as amended, sections 42 U.S.C. 1862(a), 1863(j)(1), and 1885d;
 - NSF 58 - National Science Foundation Act of 1950, as amended, 42 U.S.C. 1862(a), 1863(j)(1), and 1885d.
- g. Is this an Exhibit 300 system/project? '
- No. '

3. Data in the System

- a. What data is to be collected?
- Under NSF-43 – Doctorate Work History File: Individual names and demographic, educational, and professional characteristics of doctorate holders. Included are such parameters as age, race, geographic location, earned degrees, major subject of degree, employment status, field of employment, type of employer, primary work activity, and salary.
 - The SESTAT Data Tool imports and uses NSF-43 data; SESTAT does not include an individual's name or other direct identifiers.
- Under NSF-58 – National Survey of Recent College Graduates and Follow-Up File: Educational, professional, and demographic characteristics of degree holders including name, age, race, ethnicity, sex, disability, country of birth, social security number, occupational information, labor force status, professional activities, academic degrees, earlier education, continuing education, marital status, spouse's employment status, number and ages of children living at home, parent's educational attainment, citizenship.
 - The SESTAT Data Tool imports and uses NSF-58 data; SESTAT does not include personally identifiable information such as an individual's name or Social Security Number (SSN).
- b. What are the sources of the data? '
- National Survey of College Graduates (NSCG) - conducted for National Science Foundation by U.S. Department of Commerce, Economics and Statistics Administration, US Census Bureau.

-
- National Survey of Recent College Graduates (NSRCG) –conducted for NSF by Mathematica Policy Research, Inc. in 2008, 2010.
 - Survey of Doctorate Recipients (SDR) - conducted for NSF by National Opinion Research Center at University of Chicago.
- c. Why is the data being collected? '
- SESTAT data is collected to provide information about scientific and technical personnel in the United States, in support of NSF and US Federal and state government policy and funding decisions.
- d. What technologies will be used to collect the data?
- SESTAT data collection methodologies include; mail survey form, web survey form, and CATI (computer assisted telephone interview).
- e. Does a personal identifier retrieve the data?
- No, personal identifiers are removed from survey data before it is assimilated into the NSF SESTAT data. The SESTAT Data Tool query capability does not include use of personal identifier parameters.
 - Internally, individual SESTAT survey data records are identified by unique 9 character alpha-numeric reference ID (REFID) codes which are not available to the public. The first 2 characters of the REFID typically identify the associated survey and year of an individual record and the remaining characters are assigned by the individual survey data collection contractor. REFID codes do not contain personally identifiable information.

4. Attributes of the Data (use and accuracy)

- a. Describe the uses of the data:
- Congress has directed NSF to provide “a central clearinghouse for the collection, interpretation, and analysis of data on scientific and engineering resources and to provide a source of information for policy formulation by other agencies of the Federal Government....”
 - Data is aggregated for use in analyzing current trends in science and technology education and research.
- b. Does the system analyze data to assist users in identifying previously unknown areas of note, concern or pattern?
- Yes, data is analyzed by SBE/National Center for Science and Engineering Statistics (NCSES) personnel and is presented in publications as summary survey findings.
 - Data can be analyzed by SESTAT users through the SESTAT Data Tool.
- c. How will the data collected from individuals or derived by the system be checked for accuracy?

- Organizations such as the US Census Bureau, Mathematica Policy Research, and National Opinion Research Center of the University of Chicago conduct survey data collections comprising the SESTAT data.
- NSF contract specifications include provisions that assure data accuracy. There are three phases of survey data collection, and each phase has mechanisms to assure data accuracy as follows:
 1. The initial phase is data collection via paper questionnaire, phone survey, or web survey. Organizations have contracts with NCSES to perform data collection, and are bound to contract specifications.
 2. The second phase is data integration where inaccurate and irrelevant parts of the data files are identified and then replaced, modified or deleted as required. Organizations have contracts with NCSES to perform data integration, and are bound to contract specifications.
 3. The third phase is quality checking where data is validated. Individuals or organizations have contracts with NCSES to perform quality checking, and are bound to contract specifications.

5. Sharing Practices

- a. Will the data be shared with any internal or external organizations?
 - The public-use data is available through the NSF website.
 - Private individuals and non-governmental organizations use SESTAT data as a resource of summary statistical information about science education and professions.
 - The US Federal and state government agencies use SESTAT data for developing policy and funding programs related to scientific and technical education and research.
- b. How is the data transmitted or disclosed to the internal or external organization?
 - The SESTAT users can access public-use data via the SESTAT Data Tool; links to completed publications; and a link for submitting a download request for a SESTAT public-use data file.
 - In certain cases where the protection of respondent confidentiality would require extensive recoding that the resultant file would have little analytical utility, NSF and a qualified researcher and the researcher's institution may execute a license agreement that permits the researcher to use restricted-use SEATAT data files under strict terms and conditions.
 - NCSES employee and contractor use of restricted-use SESTAT data is governed by a defined set of access privileges based upon their assigned duties.
- c. How is the shared data secured by external recipients?
 - Public-use data does not need to be secured.
 - Licensees for restricted-use SESTAT data files are bound by the terms and conditions of the license which specifies data security requirements that must be met and maintained by the Licensees.
 - US Federal government agencies are subject to the Privacy Act of 1974 and the E-Government Act of 2002.

6. Notice to Individuals to Decline/Consent Use

- Was notice provided to the different individuals prior to collection of data?
 - SESTAT questionnaires include a Privacy Act notice and an OMB notice 3145-0019, which states the specific uses of the data collection. Survey questionnaires may be found at the following link:
<http://www.nsf.gov/statistics/question.cfm#ScienceandEngineeringWorkforce>
- Do individuals have the opportunity and/or right to decline to provide data? '
 - Yes. '
- Do individuals have the right to consent to particular uses of the data? '
 - Yes. '

7. Access to Data (administrative and technical controls)

- a. Is the data secured in accordance with FISMA requirements? '
 - Yes. '
- b. Which user group(s) will have access to the system?
 - Public and Restricted users have a defined set of access privileges within the SESTAT Data Tool.
- c. How is the access to the data by a user determined? Are procedures documented?
 - Each class of SESTAT users has a defined set of privileges and data access.
 - The Public class of users does not need logon credentials (User ID and Password), and do not sign Terms and Condition statements. Public users have access to all data that is not restricted.
 - Restricted classes of users are required to log on with a User ID and Password. Restricted classes of users are required to accept NSF warning notices when logging on to the NSF website. Restricted data users are limited to privileges and data in each assigned class.
 - Access and use of restricted-use data is determined in accordance with the [NSF/NCSES Restricted-Use Data Procedures Guide](#).
- d. How are the actual assignments of roles and rules verified according to the established security and auditing procedures?
 - The SESTAT Data Tool tracks the access of restricted classes of users. An audit record is kept of an individual's User ID and Password, date, time and of SESTAT data accessed.
- e. What auditing measures/controls and technical safeguards are in place to prevent misuse (e.g., unauthorized browsing) of data?
 - SESTAT data access is restricted to the seven classes of users where a specific user class has a defined set of permissions on the SESTAT Data Tool.

- There are two versions of the SESTAT Data Tool, found on the SESTAT home page, as follows: Public users can access only a limited set of variables. Restricted data users have access to the full set of report variables.
 - SESTAT data are provided to a user through a publication or through the SESTAT Data Tool. The SESTAT Data Tool has no browsing capability.
- f. Describe the privacy training provided to users, either generally or specifically relevant to the program or system?
- NCSES website provides access to the NSF Privacy Policy via a Privacy Policy link.
 - Restricted data users, who are NSF employees and contractors, receive training on the Privacy Act of 1974 and the E-Government Act of 2002 through annual Security Awareness Training. IT Security Awareness Training provides the protection requirements for NSF sensitive information; the various Federal laws and guidance that relate to the protection of privacy in individuals and business; and an introduction to NSF's privacy policies.
 - SESTAT Data Tool has a banner on the Terms and Conditions of use.
- g. Will NSF contractors have access to the system? If so, will they be trained on privacy principles?
- Yes, contractors may be assigned to a SESTAT user class, and are limited to the permissions of the assigned class.
 - NSF employees as well as NSF contractors are trained in privacy principles.
- h. Has the retention schedule been established by records management? If so, what is the retention period for the data in the system?
- Yes, data in the SESTAT Data Tool are retained indefinitely.
- i. What are the procedures for identification and disposition of the data at the end of the retention period?
- Data in the SESTAT Data Tool are retained indefinitely.

8. Privacy Analysis

Given the amount and type of data being collected, discuss what privacy risks were identified and how they were mitigated.

- Data Minimization: The SESTAT Data Tool provides access to survey data which has been cleared of personally identifiable information. The SESTAT Data Tool has privileged classes of users who have different levels of access; the restricted classes of data include personally identifiable information. NSF/NCSES requires a license from researchers who want access to the restricted data; licensed access to restricted SESTAT data does not include access to personally identifiable information. Individual researchers must apply for a license through an organization (e.g., a university, a research institution). For additional information on SRS data licenses, see <http://www.nsf.gov/statistics/license/start.cfm>.
- Transparency: Data collected for SESTAT system is done under contract with external survey organizations. These contracts include requirements to notify

survey respondents of the intended uses of the data. NSF personnel are notified upon logon to the NSF network that they are subject to the privacy laws of the US Federal Government.

- Accuracy: Data collected and imported into the SESTAT Data Tool goes through a three-phased system, and each separate phase includes mechanisms to insure data accuracy. Inaccurate and irrelevant parts of the data files are identified and then replaced, modified or deleted as required.
- Security: The SESTAT Data Tool provides two basic levels of access and security: Public and Restricted. Public access does not require logon credentials such as User ID and Password, and also does not require Terms and Conditions agreements or US Federal Government warning notices.
 - Restricted access is limited to National Center for Science and Engineering Statistics staff and does require logon credentials, Terms and Conditions agreements, US Federal Government warning notices, and training in meeting US Federal Government privacy laws.
 - The SESTAT Data Tool web page has a Warning Notice; Restricted users are required to Accept or Decline to accept the Warning Notice while using the SESTAT Data Tool web page.