

SF-83 SUPPORTING STATEMENT

for

Survey of Earned Doctorates

2020 and 2021 Survey Cycles

Section A

TABLE OF CONTENTS

SECTION A: JUSTIFICATION.....	3
A.1. Necessity for Information Collection.....	3
A.2. Uses of Information.....	4
Current Uses of the SED at the Federal Level.....	5
Academic and Other Uses of the SED.....	7
A.3. Consideration of Using Improved Technology.....	9
A.4. Efforts to Identify Duplication.....	10
A.5. Efforts to Minimize Burden on Small Business.....	10
A.6. Consequences of Less Frequent Data Collection.....	11
A.7. Special Circumstances.....	11
A.8. Federal Register Announcement and Consultations Outside the Agency.....	11
Consultations outside the Agency.....	11
A.9. Payment or Gifts to Respondents.....	13
A.10. Assurance of Confidentiality.....	13
A.11. Justification for Sensitive Questions.....	14
A.12. Estimate of Respondent Burden.....	14
A.13. Cost Burden to Respondents.....	16
A.14. Cost to the Federal Government.....	16
A.15. Program Changes or Adjustments.....	16
A.16. Tabulation and Publication Plans and Project Schedule.....	16
Project Schedule.....	17
A.17. Display of OMB Expiration Date.....	17
A.18. Exception to the Certification Statement.....	17

LIST OF ATTACHMENTS

- Attachment 1: 2020 SED Questionnaire
- Attachment 2: Current Representatives from Sponsoring Agencies
- Attachment 3: Authorizing Legislation of Sponsoring Agencies
- Attachment 4: Example of SED Institutional Profile
- Attachment 5: First Federal Register Announcement
- Attachment 6: NSF Staff and Contractor Data Use Agreement for Individuals
- Attachment 7: SED Institution Communication Materials
 - Job Aids for Institution Contacts
 - Graduation List Template
 - Address Roster Form
 - Missing Information Roster
- Attachment 8: SED Nonrespondent Communication Materials
- Attachment 9: SED Methodological Research

SECTION A: JUSTIFICATION

This request is for an OMB clearance for two years covering the 2020 and 2021 cycles of the Survey of Earned Doctorates (SED). The request represents an extension of a currently approved data collection (OMB No. 3145-0019). No changes to the current SED instrument or methodology are planned in the 2020 cycle. For the 2021 cycle, the National Center for Science and Engineering Statistics (NCSES) plans to review the SED questionnaire to improve the education history questions. In addition, NCSES plans to update the field of study taxonomy to include newly emerging fields and to align to the NCSES taxonomy of disciplines for the 2021 cycle. The pretest plans and any subsequent questionnaire changes will be submitted for OMB approval at a later time.

A.1. Necessity for Information Collection

The SED is sponsored by the NCSES within the National Science Foundation (NSF) in collaboration with the National Institutes of Health (NIH), the U.S. Department of Education (ED), and the National Endowment for the Humanities (NEH). Sponsoring agencies typically provide funding for the SED, obtain customized tabulations from the survey, and receive related reports. The representatives of each sponsoring agency are listed in Attachment 2. The participating federal agencies are subject to change, pending funding availability. The NCSES has lead responsibility for the SED and RTI International serves as the SED data collection contractor on behalf of NCSES through a competitively awarded procurement that covers survey operations through 2022 SED survey cycle.

The authority to collect information for the SED is established under the National Science Foundation Act of 1950, as amended, Public Law 507 (42 U.S.C. 1862), Section 3(a) (6), which directs the 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 formation by other agencies of the federal government..." Furthermore, Executive Order 10521 (March 17, 1954) states: "The Foundation shall continue to make comprehensive studies and recommendations regarding the Nation's scientific research effort and its resources for scientific activities, including facilities and scientific personnel, and its foreseeable scientific needs, with particular attention to the extent of the federal government's activities and the resulting effects upon trained scientific personnel." More recently, NCSES was established within the National Science Foundation by Section 505 of the America COMPETES Reauthorization Act of 2010 and given a broader mandate to collect data related to STEM education, the science and engineering workforce, and U.S. competitiveness in science, engineering, technology, and Research & Development.

The other Federal collaborating agencies also have statutory authority for the collection of information relevant to their mission. The following is a list of the applicable legislation:

1. NIH: Title I of the National Research Act of 1974 (PL 93 348);
2. ED: Education Sciences Reform Act of 2002;
3. NEH: Section 956(k) of the Arts, Humanities, and Museums Amendments of 1990, as enacted in Public Law 101-512.

Attachment 3 provides the cited legislation for each collaborating agency.

The SED began in academic year (AY) 1958 to collect data annually on the number and characteristics of individuals receiving research doctoral degrees from accredited U.S. institutions. Since then, all individuals receiving such doctorates are asked to complete the survey. A research doctorate is a doctoral degree that (1) requires the completion of an original intellectual contribution in the form of a dissertation or an equivalent culminating project (e.g., musical composition) and (2) is not primarily intended as a degree for the practice of a profession. The most common research doctorate degree is the PhD; in 2017, 98.0% of research doctorates awarded were PhDs. Doctorate recipients of professional doctorate degrees such as MD, DDS, JD, PharmD, and PsyD are not included in the survey, unless they also received a research doctorate.

The instrument is designed to collect information about recent doctorates education histories, funding sources, and postdoctoral plans. (Attachment 1 contains the 2020 questionnaire, which is unchanged in content and format from the 2019 questionnaire.) The results of this annual survey are used to assess characteristics and trends in research doctorate education and degrees. This information is vital for education and labor force planners within the federal government and in academia.

The SED is also used to identify sample members for NCSES's Survey of Doctorate Recipients (SDR). The SDR is designed to provide demographic and career history information about a sample of individuals with doctoral degrees in science, engineering and health (SEH) fields. Contact information obtained by the SED is used for locating the recently awarded doctorate recipients, who are added to the SDR sample every two years. The SDR results are used by all sectors (education, industry, and government) to understand trends in employment and salaries for doctorate holders in SEH fields. Results are also used to evaluate the effectiveness of equal opportunity efforts. Additionally, the results are important for internal planning because most NSF grants and fellowships are awarded to individuals who have earned, or are in the process of earning, doctoral degrees.

A.2. Uses of Information

The SED is an accurate, timely source of information on one of our nation's most precious resources – individuals with research doctorates. The SED uniquely provides comprehensive information on the educational history and early career commitments of recent U.S.-educated doctorate recipients. The resulting information is a valuable resource for government agencies, universities, professional societies, academic researchers, policymakers, program evaluators, and individuals doing research in science policy, graduate education, economics, and human resource planning.

Each academic year, the results of the SED become part of the Doctorate Records File (DRF), a complete database of more than 2 million U.S.-educated doctorate recipients from 1920 to the present.

The collaborating agencies have made extensive use of the SED. Detailed tables, tabulations, and data are used by these agencies in program planning and evaluation, policy development, and

dissemination. Similarly, detailed tables and data files are available to the doctorate-granting institutions that participate in the SED for their doctorate recipients.

There is no public-use SED data file available; however, selected SED data items are available to the public via web through the NCSES Interactive Data Tool. NCSES publishes detailed statistical tables each December (12 months after the close of data collection for the previous academic year), followed by additional statistical reports. In addition, organizations and individuals can request special tabulations from NCSES or the survey contractor. Lastly, researchers at U.S. institutions may gain access to the DRF by completing a NCSES Restricted-Use Data Licensing Agreement (<http://www.nsf.gov/statistics/license/>).

NCSES is also developing a Restricted Data Analysis System (RDAS) that will allow users greater access to the SED without compromising the confidentiality of respondents.

Current Uses of the SED at the Federal Level

The use of SED data and reports is widespread among collaborating federal agencies and other federal organizations. The data are used for policy development, program administration, and program evaluation. Some of the more important recent uses, organized by agency, are listed below.

a. The National Science Foundation

NCSES has conducted the SED since AY 1958. Special survey data tabulations constitute a key resource in meeting NCSES and NSF policy and program needs. Examples of SED uses include the following:

- NSF's Congressionally-mandated biennial reports, *Science and Engineering Indicators* (<https://www.nsf.gov/statistics/indicators>), and *Women, Minorities, and Persons with Disabilities in Science and Engineering* (<https://www.nsf.gov/statistics/wmpd>).
- *Academic Institution Profiles* (<https://ncesdata.nsf.gov/profiles/>) and *Science and Engineering State Profiles* (<https://www.nsf.gov/statistics/states/>) websites.
- Programs within NSF, especially those dealing with women, minorities, and persons with disabilities, use data from the SED for program planning. For instance, NSF's Graduate Research Fellowship Program (GRFP) has used SED information on those who complete a PhD to evaluate the effectiveness of the GRFP and its design requirements. While most NSF programs focus on U.S. citizens, SED data on foreign citizens studying in the U.S. for their PhD are also useful for international comparisons and for quantifying the attraction of U.S. graduate education around the world.
- The universe frame for sample selection of doctoral scientists and engineers for NCSES's Survey of Doctorate Recipients.
- Presentations of data on doctorates awarded to minorities and women to the National Science Board and the Committee on Equal Opportunities in Science and Engineering (CEOSE) for examining the participation of these groups in graduate education.

- Improvements in support of the Careers of Doctorate Holders project, an international effort led by the Organization for Economic Co-operation and Development (OECD) of the quality and comparability of international data on doctorate holders.
- Information to outside users at the national level, including the White House Office of Science and Technology Policy and the National Academy of Sciences.
- Publication of detailed statistical tables and reports on science and engineering doctorates. The first report to be released each year is available publicly in December, 12 months following the close of data collection from the previous academic year. The most recent report, 2017, may be found at <https://nces.nsf.gov/pubs/nsf19301/>.
- Aggregation of selected variables are publicly available through the NCSES Interactive Data Tool at <https://ncesdata.nsf.gov/ids/> and the SED Tabulation Engine at <https://nces.norc.org/NSFTabEngine/>.

b. The National Institutes of Health

For more than 30 years, NIH has used the results of the SED to meet a variety of planning, evaluating, and reporting needs:

- **Planning for the medical research workforce.** NIH relies on the results of the SED to monitor PhD production in the biomedical, behavioral, and clinical sciences. This information helps NIH determine the need for investigators in these fields and, in turn, the size and distribution of its research training programs.
- **Evaluating NIH research training programs.** Because the SED has proven to be such a reliable and comprehensive source of information on new PhDs, NIH routinely uses SED results to monitor the educational outcomes of NIH predoctoral trainees and fellows and to assess its research training programs. By comparing its internal records with the results of the SED, NIH regularly monitors PhD completion rates for students participating in NIH-sponsored training programs, their time to degree, and their plans for postdoctoral study or employment. In evaluating its research training programs, NIH also uses the SED to identify comparison groups of non-NIH-sponsored students in the same fields of study.
- **Fulfilling reporting requirements.** The SED is a critical tool for Government Performance and Results Act (GPRA) reporting on the effectiveness of NIH predoctoral training grants. In addition, since 2008, NIH has used the SED results to report on the number and type of graduate degrees awarded with NIH support in its biennial report to Congress.

c. The Department of Education

ED has been a sponsor of the SED since inception in AY 1958. The Department's National Center for Education Statistics (NCES), Sample Surveys Division, provides funding for the survey and makes extensive use of a range of SED data. Reports have been published on a time-series analysis of doctorates in the field of education, as well as in other fields.

NCES has also used data on the postgraduate plans of new doctorates. Tables with trend data are annually presented in the Center's publication *Digest of Education Statistics*. NCES has also

published tables using the DRF that compare education doctorates to doctorates in other fields, by selected characteristics.

In addition to NCES, ED's programs, such as the Office of Student Financial Aid, the individual program offices, and the Office of Planning, Evaluation, and Policy Development also use SED data for evaluation purposes.

d. The National Endowment for the Humanities

The authorizing legislation for NEH tasks the Endowment to “develop a practical system of national information and data collection on the humanities, scholars, educational and cultural groups, and their audiences.” The SED meets this mandate and gives university administrators, federal funding agencies, and private foundations an annual reading of a vital index of teaching and scholarship, the national output of humanities doctorates. NEH is currently participating in an effort led by the American Academy of Arts and Sciences to develop and regularly release via the web, a set of *Humanities Indicators*. SED data on doctorate production provides a key “indicator” of the health of the humanities workforce.

e. Other Federal Agencies and Congress

Other Federal agencies have used the SED in several ways – through requests for special tabulations and tables, data files, and licensing agreements. Congressional Research Service and Congressional office staff members have contacted NCSES for information regarding several topics relevant to legislation development, such as the percent of degrees awarded to temporary visa holders and debt levels of science doctorates at graduation, and national security interests, such as nuclear engineering doctorates awarded to foreign citizens.

Academic and Other Uses of the SED

The nation's doctorate-granting institutions not only provide SED data but also use the data. Each year since 1997, NCSES has provided the dean of each graduate school a profile of their doctorate recipients' demographic characteristics, debt status, postgraduation plans, employment and other activities, compared with national and peer-institution data (see Attachment 4 for an example of an Institutional Profile).

Graduate and baccalaureate institutions use SED data for program planning, comparison with other institutions or with national figures, and in the development of affirmative action plans. The number of SED research doctorates awarded to racial/ethnic minorities by field of study is used extensively by institutions as the only reliable source on the diversity of the potential pool of applicants for academic employment positions.

Additionally, doctorate-granting institutions participating in the SED may request cumulative microdata files going back to 1920 and preliminary data for the current academic year of their own institution's doctorate recipients. During the 2017 and 2018 survey cycles, NCSES responded to over 100 requests for data by graduate deans, other academic administrative offices, and individual researchers. Requests were made for institutional datasets, bachelor's and master's degree origins data, and preliminary institution data files. Universities help administer

the SED, and in return they have access to their data; it is a mutually beneficial data collection effort.

Researchers can apply for access to selected SED microdata under the NCSES Restricted-Use Data Licensing Agreement (<https://www.nsf.gov/statistics/license/>) if publicly available data do not address their specific needs. The NCSES Licensing Agreement, executed between an institution and NCSES, requires implementing stringent security procedures to ensure the protection of confidential data against unlawful disclosure.

Some of the recent research published using the SED data are as follows:

- Langin, Katie. “For new PhD recipients, postgraduation employment numbers are on the rise.” *Science*, December 2018
- Jackson, Joanna R., et al. “Graduation and Academic Placement of Underrepresented Racial/Ethnic Minority Doctoral Recipients in Public Health Disciplines.” *Public Health Reports*, November 2018
- Deen, Lango. “Investing in STEM and Scholarship.” *U.S. Black Engineers*, October 2018
- Kahn, Shulamit, MacGarvie, Megan. “Immigration Policy and Stay Rates of STEM PhDs.”, *International Conference on Science and Technology Indicators (STI 2018)*, September 2018
- “America’s Top Colleges 2018.” *Forbes*, August 2018
- Brown-Podgorski, Brittany L., et al. “Employment Trends Among Public Health Doctoral Recipients.” *Research and Practice*, July 2018
- Goldstone, Heather, Partan, Elsa. “Lack of Progress on Diversity in the Sciences.” *Cape, Coast, and Islands NPR*, July 2018
- Kniffin, Kevin M., Hanks, Andrew S. “The Trade-Offs of Teamwork Among STEM Doctoral Graduates.” *American Psychologist*, May 2018
- Bernard, Rachel E., Cooperdock, Emily H.G. “No Progress on Diversity in 40 Years.” *Nature Geoscience*, May 2018
- Jones, Todd R., Ehrenberg, Ronald G. “Are High-Quality PhD Programs at Universities Associated with More Undergraduate Students Pursuing PhD Study?” *Cornell Higher Education Research Institute*, April 2018
- Flaherty, Colleen. “How New Humanities PhDs Fare.” *Inside Higher Ed*, June 2018; Redden, Elizabeth. “Will U.S. Restrict Visas for Chinese Students.” *Inside Higher Ed*, March 2018
- Reys, Robert. “Doctoral Programs in Mathematics Education: Should Programs be Accredited?” *Journal of Mathematics Education*, Spring 2018
- Baum, Sandy, Steele, Patricia. “After Graduate and Professional School: How Students Fare in the Labor Market.” *Urban Institute*, February 2018
- ScienceDirect, Audrey Jaeger and Alessandra J. Dinin, “The Postdoc Landscape: The Invisible Scholars”, October 2017

A.3. Consideration of Using Improved Technology

Planning for a web-based SED started in late 1999 and has been implemented, refined, and expanded since that time. The web-based survey eases the burden on students, helps to ensure continued high response rates, improves data quality through built-in quality control processes, and provides a convenient option for institutions that publish graduation instructions online.

Offering a web version is appealing to many students. It is also very practical for respondents who are relocating at the close of their studies and are not near the graduate offices for submission of completed paper questionnaires. The web-based system has been widely accepted by both graduates and institutions. For the 2017 cycle, 94.9% of surveys were completed via the web. This demonstrates a continued increase from 2015 and 2016, when web completes represented 93% and 94% of total survey completions, respectively. Significant efforts have been made to increase schools' web participation. As of the 2019 cycle, only 7 institutions distribute and submit the paper questionnaire as their primary mode of survey completion.

The SED's electronic procedures are focused on three components: a web survey (with a parallel paper questionnaire); nonrespondent follow-up via email; and a web interface used by institutions for survey administration purposes. When a student applies for graduation, the Institutional Coordinator (IC) at their university provides them with the link to the survey registration website. The student then accesses the web survey directly after registering. Upon registering, students are sent an email containing a PIN and temporary password. For security purposes, students are prompted to create a password of their choosing upon entering the survey. This PIN and password may be used to complete the survey at a later time, including after graduation.

The IC, typically located in the graduate dean's office, is the main SED interface with the doctorate recipient. In addition to administering the survey, ICs track the status of student surveys and submit graduation lists to the SED survey contractor via the Institution Contact Administrative Tool (ICAT) on the web. At institutions where the List Coordinator (LC) role is assigned, the LC typically submits the graduation lists and other documentation to the SED survey contractor. The SED survey contractor works with the IC and LC to adjust administrative tasks as needed, to fit with each institution's procedures for processing and awarding doctoral degrees.

The ICAT allows for two levels of interaction: Level 1, a publicly accessible webpage, which includes general information and communications, such as important survey dates, general SED informational materials, and data products, and Level 2, which is school-specific, requires a username and password to access. Once logged in, ICs and LCs can monitor survey completion status of their graduates, upload graduation lists, review the graduation roster, and track the response rate for their institution or school. ICs and LCs have equivalent access privileges in the ICAT.

An additional benefit of the web option is that institutions can link the SED survey to their institution-specific exit survey, a seamless transition for students. This feature reduces the students' and institutions' burden.

A.4. Efforts to Identify Duplication

During collaborations with other agencies and organizations, NCSES has confirmed that no other government survey gathers identical information to the SED. NCSES also learns about other survey efforts and potential duplication through contacts with professional societies and groups (such as the Council of Graduate Schools and Association for Institutional Research) within both the higher education and data collection communities.

SED survey content is coordinated with NCSES's Survey of Doctorate Recipients (SDR), Early Career Doctorates Survey (ECDS), and National Survey of College Graduates to assure relevant uniform approaches on similar items such as race/ethnicity and specific functional limitations.

The Integrated Postsecondary Education Data System (IPEDS) also provides information on doctoral degrees. Differences between the SED and IPEDS are outlined below. While SED collects data from individuals, the IPEDS Completions survey, conducted by the National Center for Education Statistics (NCES) of the Department of Education, collects aggregate data from institutions on numbers of degrees at each level by discipline and on recipients (including race/ethnicity and sex), while the SED collects from individual doctorate recipients information on more than two dozen variables, not collected by the IPEDS survey and used by NCES.

There are four duplicative data items collected on both the SED and IPEDS: field of degree and the demographic variables of citizenship, sex, and race/ethnicity. However, important purposes are served by including these variables in both surveys:

- In the SED, field of degree, citizenship, sex, and race/ethnicity are frequently used in analyses with other key SED variables, such as the length of time spent pursuing the degree and the amount of debt accumulated during graduate education. The other key variables cannot be collected from the IPEDS institutions. The field of degree and demographic variables are also used to identify individuals in "rare subgroups" for oversampling in the SDR.
- IPEDS' inclusion of field of degree, citizenship, sex, and race/ethnicity permits comparative analyses of trends in degree production at different degree levels. SED data cannot be substituted for IPEDS in such comparisons because of the inevitable differences in responses from institutional and demographic surveys. For example, individuals' racial/ethnic self-identification on these variables may differ from those maintained by the institutions. Also, IPEDS collects data on types of doctoral degrees that are explicitly excluded from the SED (i.e., doctorate degrees intended for the practice of a profession, such as an MD). Hence, eliminating doctoral degrees from the IPEDS data collection would result in a loss of information about these other important types of doctoral degrees.
- Including field of degree, citizenship, sex, and race/ethnicity questions on both surveys provides important validity checks for both surveys at the aggregate level.

A.5. Efforts to Minimize Burden on Small Business

Not applicable. The SED does not collect information from small businesses.

A.6. Consequences of Less Frequent Data Collection

The SED is an important source for monitoring changes in participation in the various fields of study by demographic groups of interest (including U.S. and non-U.S. citizens on both permanent and temporary visas). The SED data pertaining to respondents' postgraduation plans provide an annual barometer of the market conditions encountered by new doctoral degree recipients and are therefore an integral component in policy implementation and program design.

Less frequent data collection would result in a more complicated survey administration by the institutions. Currently, the Institution Coordinator (IC) at each institution requests a completed survey from each person receiving a research doctorate during their final semester prior to graduation. ICs include the link to the web survey with other electronic documents related to graduation or insert the SED questionnaire into the package of materials for doctorate recipients. Any less frequent collection of the SED would yield far lower response rates because the graduate deans' offices would be uncertain about the timing and distribution of questionnaires to prospective doctoral graduates, a process which now occurs continuously throughout the survey year. Discussions with the Council of Graduate Schools and several universities confirm that graduate schools would face extreme difficulty if the survey were operated on a non-annual basis. Stability of both the survey questionnaire and the survey collection process is imperative for data usefulness and ease of administration.

If the SED were conducted less frequently, there would also be significant repercussions for the SDR sample selection. Locating information obtained from the SED is necessary for contacting the new research doctorate recipients who are added to the SDR sample. The coordination of timing, content, and procedures of these two studies is critical to the success of both the SED and SDR.

A.7. Special Circumstances

Not applicable. This data collection does not require any of the reporting requirements listed.

A.8. Federal Register Announcement and Consultations Outside the Agency

The Federal Register announcement for the SED appeared on October 29, 2018 (see Attachment 5).

Consultations outside the Agency

NCSES often invites others to comment on the SED. Comments have come from the SED collaborating federal agencies, expert panels convened by NCSES, the Council of Graduate Schools, and other agencies and academic institutions. NCSES has also received comments from respondents, university faculty advisors, graduate deans' offices, and professional researchers by telephone, email, mail, and in-person contacts. NCSES seeks input from university representatives at venues such as professional conferences, meetings and personal site visits to institutions. These consultations have identified problems with survey administration or in the interpretation of certain data items. NCSES often discusses identified issues with ICs for

their conceptual validity and applicability to all fields of study, and assesses next steps based on needs and respondent and institutional burden.

The collaborating agencies meet periodically to discuss the SED design, operation and dissemination activities, and to plan future activities. They review recent trends in the number of doctorate recipients receiving degrees in emerging fields of study – that is, fields of study not currently coded within the SED taxonomy – and in fields of study for which there are few graduates. This review is the basis for the decisions made every two years on SED taxonomy changes.

NCSES has convened multiple meetings of a Human Resources Expert Panel (HREP) to improve data collected on the education and employment of the science and engineering (S&E) labor force through review and renewal of the program’s surveys, and to promote use of the data for research and policy analysis purposes. HREP accomplishes its mission by: 1) suggesting methods to publicize and promote the data; 2) providing advice on efforts to improve the timeliness and accuracy of S&E education and labor force data; 3) providing a mechanism for obtaining ongoing input from the community interested in S&E personnel data; 4) providing perspectives on the data needs of decision makers; 5) identifying issues and trends that are important for maintaining the relevance of the data; and 6) proposing ways to enhance the content of the NCSES education and workforce surveys. The panel consists of a rotating membership of between 12 and 15 individuals who represent the sciences, academia, business/industry, government, researchers and policy makers.

In 2018, two expert survey methodologists, Dr. Jolene Smyth and Dr. Kristen Olson of the University of Nebraska at Lincoln, were asked to review the education history section of the SED survey instruments for potential improvement to make it easier for the respondents to provide information about their degrees. The NCSES plans to submit any SED questionnaire changes for OMB approval at a later time.

The National Academies of Science, Engineering, and Medicine’s Committee on National Statistics (CNSTAT), at the request of NCSES, convened an expert panel to review, assess, and provide guidance on NCSES’s effort to measure the S&E workforce population in the United States. Given the evolving data needs of NCSES stakeholders and the budget climate uncertainty under which NCSES operates, NCSES would like to develop a framework for measuring the S&E workforce that will enable the flexibility to examine emerging issues related to this unique population while at the same time allowing for stability in the estimation of trend data. This framework would provide direction for numerous issues related to measuring the S&E workforce population including content, data sources, survey design and methodology, data collection, data processing, data integration, data dissemination, and data promotion.

At the end of its review, the panel issued a report with findings, recommendations, and priorities for improving the relevance, accuracy, timeliness, and cost-effectiveness of S&E workforce data for the next decade and beyond.¹ While discussing the SED data collection and processing effort, the CNSTAT panel noted “Although maintaining these response rates is becoming increasingly difficult as part of a trend that affects all surveys, the [SED] data collection procedures appear well suited to maximizing response rates efficiently.”

¹ <https://www.nap.edu/catalog/24968/measuring-the-21st-century-science-and-engineering-workforce-population-evolving>

NCSES conducts institution site visits focused primarily on improving poor response rates and resolving data collection problems. The site visits also allow for a discussion of the uses of SED data. Since the last clearance, site visits have been made to four institutions (Arizona State University, University of Texas-Austin, North Carolina State University, and Duke University).

Other Consultations

NCSES has numerous other contacts with the user community, including staff of organizations such as the National Postdoctoral Association, the Association of American Medical Colleges, the Association for Institutional Research, the Council of Graduate Schools, the American Association of Universities, and the Organization for Economic Co-operation and Development. Routine information requests also provide insight into the interests of the public.

A.9. Payment or Gifts to Respondents

No incentives in the form of payment or gifts to the doctoral graduates are used in the SED.

A.10. Assurance of Confidentiality

The SED is collected in conformance with the strict confidentiality requirements found in the NSF Act of 1950, as amended. The SED is also collected in conformance with the Privacy Act of 1974, including the section of the Privacy Act requiring notification of the respondent concerning the data uses and the voluntary nature of their responses. The confidentiality pledge to SED respondents follows.

This information is solicited under the authority of the National Science Foundation Act of 1950, as amended. All information you provide is protected under the NSF Act and the Privacy Act of 1974 and will be used only for research or statistical purposes by your doctoral institution, the survey sponsors, their contractors, and collaborating researchers for the purpose of analyzing data, preparing scientific reports and articles, and selecting samples for a limited number of carefully defined follow-up studies. Per the Federal Cybersecurity Enhancement Act of 2015, your data are protected from cybersecurity risks through screening of the federal information systems that transmit your data. The last four digits of your Social Security Number are also solicited under the NSF Act of 1950, as amended; provision of it is voluntary. It will be kept confidential. It is used for quality control, to assure that we identify the correct persons, especially when data are used for statistical purposes in Federal program evaluation. Any information publicly released (such as statistical summaries) will be in a form that does not personally identify you or other respondents. Your response is voluntary and failure to provide some or all of the requested information will not in any way adversely affect you.

The time needed to complete this form varies according to individual circumstances, but the average time is estimated to be 20 minutes. If you have comments regarding this time estimate, you may write to the National Science Foundation, 2415 Eisenhower Avenue, Alexandria, Virginia 22314, Attention: NSF Reports Clearance Officer. A Federal agency may not conduct or sponsor a collection of information unless it displays a currently valid OMB control number.

Specific procedures for protecting both hard copy and electronic data are used by the survey contractor. All project staff are required to sign confidentiality agreements before they first access any SED data, and on a yearly basis thereafter (see Attachment 6). Data files with personal identifiers are provided to NIH and its contractors, a collaborating federal agency, through an NCSES Restricted-Use Data Licensing agreement, with all contractors signing data use agreements. As indicated explicitly in the confidentiality statement, the graduate dean of the respondent’s institution may request data for respondents from that institution only with a written agreement to use such data for statistical and program evaluation purposes only. Lastly, SED data files with personal identifiers are provided to the SDR contractor under a signed data use agreement, to locate the SDR’s selected sample members. No one outside of these groups can obtain data files with direct identifiers such as email addresses, phone numbers and mailing addresses. Qualifying researchers can obtain SED microdata, without direct identifiers, only by executing a Restricted-Use Data Licensing Agreement with NCSES.

A.11. Justification for Sensitive Questions

The SED recognizes the growing sensitivity towards requesting respondents’ Social Security number. The SED is provided authority to collect respondent Social Security numbers under the NSF Act of 1950 (42 U.S.C. 1861 et seq.), as amended, and in accordance with the Privacy Act of 1974. However, the SED collects only the last four digits of the Social Security number to be used to ascertain the correct identity of survey respondents for survey operations and evaluation purposes, and to match SED data to data related to federal funding support for graduate education.

A.12. Estimate of Respondent Burden

The SED is a census of all individuals receiving a research doctorate in the United States in an academic year. In 2020, an estimated 56,522 individuals are expected to receive research doctorates from U.S. institutions. Using the target response rate of 91%, the number of SED respondents in 2020 is estimated to be 51,435 (56,522 estimated doctorate recipients × 0.91). Similarly, the number of individuals expected to earn research doctorates in 2021 is estimated to be 57,006; the number of respondents is estimated to be 51,875 (57,006 estimated doctorate recipients × 0.91).

The average response time for the 2020 and 2021 questionnaires is estimated to be 20 minutes. Thus, the total respondent burden for the 2020 questionnaire is estimated to be 14,402 hours (51,435 respondents × 20 minutes) and 14,525 hours for the 2021 SED questionnaire (53,100 respondents × 20 minutes).

Year	Doctorates	Respondents (91%)	Minutes per case	Total Minutes	Total Hours	Cost @\$30/hour
2020	56,522	51,435	20	1,028,700	17,145	514,350
2021	57,006	51,875	20	1,037,500	17,292	518,750

The estimated cost to respondents for the 2020 data collection is \$514,350. This is based on the estimated 17,145 hours of response burden at a time-cost of \$30.00 per hour. Based on an estimated 17,292 hours of response burden and a time-cost of \$30.00, the estimated cost to respondents for the 2021 data collection is \$518,750. The \$30.00 per hour time-cost estimate is derived from the results of the 2017 SED, which indicate that the median income for doctorate recipients with known employment or postdoc commitments is \$60,000. Assuming a 40-hour work week and 50 weeks of work per year, an annual salary of \$60,000 equates to \$30.00 per hour.

In addition to having students complete the SED, NCSES also requires that institutions collect administrative data. The IC at each school or institution distributes the survey registration URL, monitors survey completion status, and submits graduation lists to the SED survey contractor. If using paper questionnaires, the IC returns these via mail. To fulfill their role, ICs must complete the following forms (see Attachment 7 for examples of the materials):

- Graduation List requests the names, fields, and contact information for eligible graduates
- Address Roster Form requests a physical mailing address, phone number, and/or e-mail address for nonrespondents (requested only if information has not been supplied on the Graduation List)
- Missing Information Roster requests critical items for nonrespondents. This includes birthdate, sex, citizenship, race, ethnicity, bachelor’s institution, doctoral field of study, and post-degree location.

Based on focus groups conducted with Institution Contacts (ICs), it is estimated that the SED requires no more than 1% of the IC’s time, which computes to 20 hours per year (40 hours per week x 50 weeks per year x .01). At an estimated median hourly wage rate of \$27.78 for approximately 617 ICs (representing 617 schools from 450 institutions), the total estimated time-cost of administering the SED in 2020 is \$342,805 per year. With a projected 620 ICs in 2021, the estimated time-cost of administering the SED in 2021 is \$344,472. The \$27.78 median hourly wage estimate is derived from the Bureau of Labor Statistics’ “May 2017 National Occupational Employment and Wage Estimates”

(https://www.bls.gov/oes/current/oes_nat.htm#43-0000) by including a combination of the Office and Administrative Support Occupations (\$16.70 median hourly wage, representing 60% of ICs) and Education Administrators, Postsecondary (\$44.41 median hourly wage, representing 40% of ICs) figures.

The chart below summarizes the annual burden anticipated for all the tasks involved with conducting the SED:

Description	# of Responses*	Respondent Burden	Annual Burden Hours	Average Hourly Wage Rate	Annual Cost Burden
Doctorate recipients completing SED	51,655	20 minutes	17,218	\$30.00	\$516,550
Institutional Contacts administering SED	619	20 hours	12,380	\$27.78	\$343,916
Total SED Annual Burden	52,274		29,598		\$860,466

* Mean of estimated respondents in 2020 & 2021 survey rounds

A.13. Cost Burden to Respondents

Respondents need not purchase, operate, or maintain capital equipment, software, or storage facilities. There is no actual cost to the SED respondents other than the burden hour cost noted in A.12.

A.14. Cost to the Federal Government

The cost to the Federal Government for this annual data collection is approximately \$3.3 million per year. This amount is based on the contract cost for the 2010-2022 SED survey cycles.

A.15. Program Changes or Adjustments

The only expected cost adjustment in the 2020 and 2021 SED is from the increase in SED universe size, resulting in a higher number of respondents.

A.16. Tabulation and Publication Plans and Project Schedule

The results of the SED will be disseminated in a number of ways. To release the data, NCSES will publish a set of approximately 70 online Summary Report Data Tables. These tables will be descriptive in nature and will provide extensive information on the education and employment plans of doctoral graduates by field of study, doctorate granting institution, and demographic characteristics such as race/ethnicity, citizenship, sex, and disability. NCSES will also publish a Report Digest, with approximately 36 figures highlighting findings from key survey themes. The Digest will be available in both print and electronic formats. The printed Digest is provided to participating SED institutions and to individuals and institutions who have requested past survey results.

The SED data will also be used in the development of key NSF reports, including the Congressionally-mandated reports *Science and Engineering Indicators* and *Women, Minorities, and Persons with Disabilities in Science and Engineering*. Both of these publications, plus additional detailed tables, will be available on the NCSES website.

Aggregated data on selected SED variables are publicly available through the NCSES Interactive Data Tool (<https://ncesdata.nsf.gov/ids/>). As stated previously, NCSES is developing a Restricted Data Analysis System that will allow users greater access to the SED without compromising the confidentiality of respondents. This system, based on the NCES DataLab, which will be used to release SED data beginning with data from the SED 2018 survey cycle. Additionally, SED data will be available to licensed researchers via the NCSES's data enclave, a secure environment that provides researchers remote access to microdata while still protecting respondent confidentiality.

The Scientists and Engineers Statistical Data System (SESTAT), available via NCSES's website, includes some SED variables that pertain to SDR respondents (e.g., past educational degrees). The SESTAT system provides a rich online resource for producing tabulations. As noted above, microdata are also disseminated to collaborating agencies and licensed researchers. Their results are expected to generate reports and other publications that further disseminate the data. Finally, it is anticipated that substantive SED data analyses will be presented at relevant professional

meetings, such as the annual meetings of the Association for Institutional Research, the Council of Graduate Schools, the American Educational Research Association, the American Statistical Association, the American Economic Association, the American Sociological Association, the American Association for Public Opinion Research, etc.

The 2020 SED includes doctoral graduates from 1 July 2019 to 30 June 2020. The paper SED and corresponding materials will be mailed to the graduate schools and the web version will be uploaded upon OMB approval for continuous distribution to graduate students as they complete their doctoral requirements. Returned paper questionnaires will be edited and coded until survey close-out in December 2020 for the 2020 academic year. After the survey close-out, data variables will be constructed, edited, evaluated, and reviewed for trend consistency in January 2021. In February 2021, the file will be further evaluated, and quality control checks will be made. Data will be tabulated in April 2021 and prepared for publication by November 2021. Aggregate data will be made available to the public in December 2021 via the on-line Summary Report Data Tables on the NCSES website.

Project Schedule

The 2020 SED survey schedule follows. The 2021 SED survey schedule is expected to be identical, except lagging by one year.

Phase	Time
Receive OMB clearance approval	April 2019
Survey instrument available to students	May 2019
Data collection close-out	December 2020
Preparation of data file	February 2021
Production of publications	April 2021
Release of data by NCSES	December 2021

A.17. Display of OMB Expiration Date

The OMB Expiration Date will be displayed, as indicated.

A.18. Exception to the Certification Statement

The 2020 and 2021 SED will comply with the certification statement on form OMB 83-1.