

SF-83 SUPPORTING STATEMENT

for

Survey of Earned Doctorates

2022 and 2023 Survey Cycles

Section A

TABLE OF CONTENTS

A.1. Why the Collection is Necessary.....	3
A.2. Uses of the Information.....	4
Current Uses of the SED at the Federal Level.....	5
Academic and Other Uses of the SED.....	7
A.3. Use of Automated, Electronic, Mechanical or Other Technological Techniques.....	10
A.4. Efforts to Identify Duplication.....	11
A.5. Impacts on Small Entities.....	12
A.6. Consequences of Less Frequent Collection.....	12
A.7. Special Circumstances.....	12
A.8. Consultations Outside the Agency.....	12
Consultations outside the Agency.....	13
A.9. Payment or Gifts to Respondents.....	15
A.10. Assurance of Confidentiality.....	16
A.11. Justification for Asking Sensitive Questions.....	16
A.12. Burden Estimate.....	17
A.13. Costs to Respondents.....	18
A.14. Costs to the Federal Government.....	18
A.15. Changes from the Prior Cycle.....	19
A.16. Plans for Tabulation or Publication.....	19
Project Schedule.....	19
A.17. Exception to Displaying of OMB Expiration Date.....	20
A.18. Exception to the Certification Statement.....	20

LIST OF ATTACHMENTS

- Attachment 1: 2021 SED Questionnaire Reference Copy
- Attachment 2: Current Representatives of the SED Sponsoring Agencies
- Attachment 3: Authorizing Legislation of the SED Sponsoring Agencies
- Attachment 4: SED Institutional Profile - Example
- Attachment 5: 2022-23 SED First Federal Register Notice
- Attachment 6: NSF NCSES Individual Data Use Agreement for NSF Staff and Contractors
- Attachment 7: SED Institution Contact Materials
- Attachment 8: SED Nonrespondent Follow-Up Contact Materials
- Attachment 9: SED Recent Methodological Research
- Attachment 10: Proposed Revisions to COVID-19 Question Module

SECTION A: JUSTIFICATION

This request is for an OMB clearance for three years covering the 2022 and 2023 cycles of the Survey of Earned Doctorates (SED). The request represents a revision of a currently approved data collection (OMB No. 3145-0019). The one change to the 2022 survey cycle involves revisions to the COVID-19 impact question items that were added to the 2021 SED to measure its impact on the doctorate recipients degree completion and postgraduation plans. Once the planned cognitive testing of the revised COVID-19 questions is completed, the final set of questions for inclusion in the 2022 and 2023 SED will be submitted for OMB approval.

A.1. Why the Collection is Necessary

The SED is sponsored by the National Center for Science and Engineering Statistics (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 the 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 2019, 98.4% 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 doctorate recipients' education histories, funding sources, and postdoctoral plans. (Attachment 1 contains a reference copy of the 2021 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 NCSSES'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 the 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 (<https://ncesdata.nsf.gov/builder/sed>). 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 an 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 data 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) routinely uses SED information on those who complete a PhD to evaluate the effectiveness of the GRFP and its design requirements.
- Participants of the NSF Engineering Research Centers (ERC) and the NSF Research Experiences for Undergraduates (REU) are linked to the SED to study the impact of these NSF programs in educational achievement and career outcomes.
- The SED data are linked to the Universities Measuring the Effects of Research on Innovation, Competitiveness, and Science (UMETRICS) data to develop data science training programs offered to staffs of NSF and other Federal agencies and the broader research communities. The project-based training offers real data experiences in exploring the role of federal funding in doctoral career outcomes.
- The universe frame for sample selection of doctoral scientists and engineers for NCSES's Survey of Doctorate Recipients (SDR).
- The SED-SDR samples are linked to bibliometrics data to produce novel research data for studying scientific output and impact of the U.S.-trained doctorate recipients, and for supporting science policy research.

- 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.
- 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 SED detailed statistical tables and reports on science and engineering doctorate recipients. 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, 2019, may be found at <https://nces.nsf.gov/pubs/nsf21308>.
- Aggregation of selected variables are publicly available through the NCSES Interactive Data Tool at <https://ncesdata.nsf.gov/builder/sed>.

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 doctorate recipients. 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 doctorate recipients to doctorate recipients 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 doctorate recipients. 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. The Congressional Research Service and Congressional office staff members have contacted NCSES for information regarding several topics relevant to developing legislation, such as the percent of degrees awarded to temporary visa holders, debt levels of science doctorate recipients at graduation, and 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 2019 and 2020 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:

Journal Articles

Diethorn, H. A. (2020). Better Safe than Sorry: The Effect of Permanent Residency Delays on the Propensity of Foreign STEM Doctorates to Work in Startups. In. Cambridge, MA: NBER.

Kinoshita, T. J., Knight, D. B., Borrego, M., & Wall Bortz, W. E. (2020). Illuminating systematic differences in no job offers for STEM doctoral recipients. *PLoS One*, 15(4), e0231567. <https://doi.org/10.1371/journal.pone.0231567>

Laird, Frank N. (2020). Sticky Policies, Dysfunctional Systems: Path Dependency and the Problems of Government Funding for Science in the United States. *Minerva*, 58:513–533 <https://doi.org/10.1007/s11024-020-09409-2>

Lebovitz, L., Swaan, P. W., & Eddington, N. D. (2020). Trends in Research and Graduate Affairs in Colleges and Schools of Pharmacy, Part 2 – Students. *American Journal of Pharmaceutical Education*. 84 (5) 7642; <https://doi.org/10.5688/ajpe7642>

Moore, N. A., & Burns, R. A. (2020). Economic Development as an Administrative Prerogative: An Event History Analysis of APLU Institutions. *Economic Development Quarterly*, 34(3), 242-268. <https://doi.org/10.1177/0891242420924458>

Randall, J., Riosb, J. A., & Junga, H. J. (2020). A Longitudinal Analysis of Doctoral Graduate Supply in the Educational Measurement Field. *Educational Measurement: Issues and Practice*. <https://doi.org/10.1111/emip.12395>

Betsey, C. L. (2019). African Americans in Economics at the University of Michigan-Ann Arbor Since the Kerner Commission Report of 1968. *The Review of Black Political Economy*. <https://doi.org/10.1177/0034644619880562>

Chang, W. Y., Cheng, W., Lane, J., & Weinberg, B. (2019). Federal funding of doctoral recipients: What can be learned from linked data. *Research Policy*. <https://doi.org/10.1016/j.respol.2019.03.001>

Fernandez, F. (2019). What's Money Got to Do with It? An Examination of the Relationships between Sources of Financial Support and the Post-Graduation Career Plans of Latina and Latino Doctoral Students in the Social Sciences. *The Review of Higher Education*, 43(1), 143-168. <https://doi.org/10.1353/rhe.2019.0093>

Krasna, H. (2019). Employment Trends Among Public Health Doctoral Graduates (letter to the editor). *Am J Public Health*, 109(4), e9. <https://doi.org/10.2105/AJPH.2019.304966>

Kulp, A. M. (2019). Parenting on the Path to the Professoriate: A Focus on Graduate Student Mothers. *Research in Higher Education*. <https://doi.org/10.1007/s11162-019-09561-z>

Reys, R., Reys, B., Shih, J., & Safi, F. (2019). EDUCATION: Doctoral programs in mathematics education: A status report of size, origin of program leadership, and recommended institutions. *Notices of the American Mathematical Society*, 66(02), 212. <https://doi.org/10.1090/noti1796>

Roach, M., & Skrentny, J. (2019). Why foreign STEM PhDs are unlikely to work for US technology startups. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1820079116>

Sula, C. A., & Hill, H. V. (2019). The early history of digital humanities: An analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004). *Digital Scholarship in the Humanities*. <https://doi.org/10.1093/lc/fqz072>

Zhang, Y., & Palma, M. A. (2019). The impact of the H-1B cap exemption on Ph.D. labor markets. *Empirical Economics*. <https://doi.org/10.1007/s00181-019-01721-5>

News Articles

Ladyzhets, B. (2020, December 16). These 6 graphs show that Black scientists are underrepresented at every level. *The Wall Street Journal*.

Flaherty, C. (2020, December 9). Calm Before the Storm. *Inside Higher Ed*.

Forde, K. (2020, October 19). Do Nobel Prizes portend women's progress in STEM fields? *Aljazeera*.

Harris, A. (2019, April 19). The Disciplines Where No Black People Earn Ph.D.s. *The Atlantic*.

Smith, N. (2019, March 27). American Employers Are Hung Up on Hiring Ph.D.s. *Bloomberg*.

Langin, K. (2019, March 12). In a first, U.S. private sector employs nearly as many Ph.D.s as schools do. *Science*.

Carey, K. (2020, March 6). The Bleak Job Landscape of Adjunctopia for Ph.D.s. *The New York Times*.

Reports and Books

Kinney, S. K., Looby, C. B., & Yu, F. (2020). Advantages of Imputation vs. Data Swapping for Statistical Disclosure Control. In *Privacy in Statistical Databases* (pp. 281-296).

National Communication Association. (2020). *A Profile of the Communication Doctorate VII: The 2018 National Science Foundation's Survey of Earned Doctorates*.

Yudkevich, M., Altbach, P. G., de Wit, H., & Publications, S. (2020). *Trends and Issues in Doctoral Education: A Global Perspective*: SAGE Publishing. India.

Zwetsloot, R., Feldgoise, J., & Dunham, J. (2020). *Trends in U.S. Intention-to-Stay Rates of International Ph.D. Graduates Across Nationality and STEM Fields*.

Hancock, S., Wakeling, P., & Chubb, J. (2019). *21st Century PhDs: Why we need better methods of tracking doctoral access, experiences and outcomes*. In London, UK: Research on Research Institute (RORI).

Hrabowski III, F. A., Rous, P. J., & Henderson, P. H. (2019). *The Empowered University: Shared Leadership, Culture Change, and Academic Success*: Johns Hopkins University Press.

Veugelers, R., & Baltensperger, M. (2019). *Europe-the Global Centre for Excellent Research Study for the Committee on Industry, Research and Energy, Policy Department for Economic, Scientific and Quality of Life Policies*.

In addition to the selected publications shown above, the SED has been used in at least 4 doctoral dissertations and 10 working papers.

A.3. Use of Automated, Electronic, Mechanical or Other Technological Techniques

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 2020 cycle, 97.3% of surveys were completed via the web. This demonstrates a continued increase from 2018 and 2019, when web completes represented 95.5% and 95.8% 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, and those institutions have agreed to distribute the web link. Thus, the paper questionnaire ceased to be used in data collection beginning with the 2020 survey cycle.

The SED's electronic procedures are focused on three components: a web survey; 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. Impacts on Small Entities

Not applicable. The SED does not impact small entities.

A.6. Consequences of Less Frequent 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 (85 FR 65078) appeared on October 14, 2020 (see Attachment 5). NCSES received one public comment from the Federation of American Societies for Experimental Biology advocating for the collection of information on the participation of sexual and gender minorities in doctoral education. NCSES informed the commenter that it shares their interest in improving federal data collections and providing reliable measures for important segments of the population. NCSES also informed the commenter that it is conducting research to evaluate these measures with the goal that this research, in combination with on-going sexual orientation and gender identity (SOGI) survey content research being conducted by other federal agencies, will enable the development of standard guidance for collecting SOGI data in the near future.

In response to the second Federal Register announcement that appeared on February 5, 2021, NCSES received two public comments. The first comment was from Dr. Roxanne Miller at Chapman University, who a) asked NCSES to examine the response options for all demographic questions to improve their quality and clarity, particularly on sex and marital status questions, and b) expressed concern for undocumented students' response to the citizenship status question. NCSES assured the commenter of its awareness of the sensitivities involved in these demographic questions, and that it is currently conducting research to examine the feasibility of collecting reliable SOGI data from our survey population. NCSES also informed the commenter that it will evaluate whether modifications to the marital status question are needed. Lastly, NCSES informed the commenter that all respondent data are kept highly confidential and will review the SED survey instrument to ensure respondents understand that they may skip questions they do not wish to answer.

The second comment was from Dr. Jon Freeman at NYU on behalf of the American Association for the Advancement of Science (AAAS) and the American Educational Research Association (AERA) urging NCSES to add sexual orientation and gender identity questions in the SED or initiate piloting of SOGI data for NCSES surveys. In response, NCSES acknowledged data users' desires for information on the prevalence of sexual and gender minorities in doctorate education and the nation's doctorate workforce. NCSES also informed the commenter of its several upcoming research efforts to test questions on sexual orientation and gender identity, using qualitative and quantitative methods, to better understand and develop new questions that can be considered for the SED and other relevant NCSES surveys.

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. Subsequently, methodological testing was conducted to determine the best format for the education history section and a revised version of that section was implemented during the 2021 SED cycle.

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 (New York University, City University of New York, University of Delaware, and Georgetown University).

Consultations Related to Collecting Sexual Orientation and Gender Identity Information

In recent years, NCSES has sought the advice and guidance of survey methodologists, statisticians, demographers, researchers, data analysts, and policymakers to examine numerous issues related to the collection of data on sexual orientation and gender identity (SOGI). NCSES is constantly striving to improve its data collection efforts to provide measures about important segments of the U.S. population, especially segments underrepresented in STEM fields. NCSES acknowledges that collecting information on the participation of sexual and gender minorities in STEM education and the nation's STEM workforce would be useful to researchers and policymakers. As such, NCSES continues to actively engage on the Federal Committee on Statistical Methodology's (FCSM) Working Group on Measuring Sexual Orientation and Gender Identity in addressing issues surrounding the collection and availability of SOGI data.

As a federal statistical agency, NCSES has taken a measured approach to adding SOGI questions to the SED and other NCSES surveys, to appropriately balance data quality, respondent burden, confidentiality, and data user needs. This spring, NCSES will be engaged in several important research efforts to better understand and develop new questions on sexual orientation and gender identity that can be considered for the SED and other relevant NCSES surveys. These efforts are outlined below.

1. NCSES will be conducting methodological research using a non-probability convenience sample of approximately 2,800 cases. The web survey instrument will include questions on sexual identity, sex at birth, current gender identity, and pronouns used by the respondents. The sexual identity question is based on the National Crime Victimization Survey and National Health Interview Survey, which are interviewer-administered surveys. In addition to assessing the performance of interviewer-administered questions in a self-administered setting, this study will include an experiment to assess possible measurement errors associated with the response options that appear with the sexual identity question. Each

respondent will receive the same question stem and one of four sets of response options. The study also includes a second experiment on the gender identity questions: all respondents will be asked the sex at birth and current gender identity questions, but the experiment will randomize the order in which the two questions are asked to measure possible context effects. Finally, the survey instrument will contain a series of follow-up probes to the SOGI and pronoun questions to address question comprehension, accuracy of response options, and sensitivity, which are typical in personal cognitive interviews.

2. To complement the non-probability-based study above, a second research effort involves collecting SOGI information from a sample of approximately 5,000 cases as part of the 2021 NSCG non-production bridge panel. The data collection conducted by the U.S. Census Bureau on behalf of NCSES, is scheduled to begin April 2021. This study will examine operational considerations related to collecting SOGI data, assess the quality of the question items through web instrument paradata, and evaluate the potential disclosure risk of SOGI estimates by examining the size of the SOGI population across various education, workforce, and demographic characteristics. The three SOGI questions that will be asked in this study are the respondent's sex assigned at birth, current gender identity, and sexual identity or orientation. The response options for the current gender identity and sexual identity questions will vary from those in the study outlined above, giving NCSES additional insights and opportunities for examination.
3. Finally, through research initiated by a Broad Agency Announcement, NCSES, in coordination with the National Institutes of Health, plans to contract a qualitative study to understand the SOGI data collected by employers across various STEM industries. The research will a) review the current state of SOGI data collection in workforce systems across sectors (including public, private, academic, and non-profit), with a focus on STEM fields, b) identify gaps in knowledge and data collection efforts, and c) highlight lessons learned and next steps for increasing SOGI data collection across STEM fields and employment sectors.

In addition to the above research efforts, additional testing may be required before adding any SOGI question to SED, because of the unique aspects of the survey and its respondents. For example, SED has a greater proportion of temporary visa holders than other federal or NCSES surveys, so additional cross-cultural testing may be needed.

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, a doctoral 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 Asking 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. Burden Estimate

The SED is a census of all individuals receiving a research doctorate in the United States in an academic year. In 2022, an estimated 57,000 individuals are expected to receive research doctorates from U.S. institutions. Using the target response rate of 92%, the number of SED respondents in 2022 is estimated to be 52,440 (57,000 estimated doctorate recipients × 0.92). Similarly, the number of individuals expected to earn research doctorates in 2023 is estimated to be 58,000; the number of respondents is estimated to be 53,360 (58,000 estimated doctorate recipients × 0.92).

The current average response time for the 2021 web survey is 20 minutes. The table below presents our estimates for respondent burden.

Year	Doctorates	Respondents	Average Response Time	Total Burden Hours	Estimated Hourly Wage	Total Cost
2022	57,000	52,440	20 minutes	17,480	\$41	\$716,680
2023	58,000	53,360	20 minutes	17,787	\$41	\$729,253

The estimated cost for the respondents to complete the 2022 SED is \$716,680. This is based on the estimated 17,480 hours of response burden at an hourly wage of \$41.00. For the 2023 data collection, based on an estimated 17,787 hours of response burden at a cost of \$41.00 per hour, the estimated cost to respondents is \$729,253. The \$41.00 per hour time-cost estimate is derived from the 2019 SED data, which indicated the median income for doctorate recipients with known employment or postdoc commitments was \$83,000. Assuming a 40-hour work week and 50 weeks of work per year, an annual salary of \$83,000 equates to \$41.00 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. To fulfill their role, ICs must complete the following forms (see Attachment 7 for examples of the institution contact 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.

Year	Institutional Contacts	Average Response Time	Total Annual Burden Hours	Estimated Hourly Wage	Total Annual Cost
2022	600	20 hours	12,000	\$29.19	\$350,280
2023	600	20 hours	12,000	\$29.19	\$350,280

The total estimated cost for the institutional contacts to administer the 2022 and 2023 SED is \$350,280 per year. 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). Based on an estimated median hourly wage rate of \$29.19 for approximately 600 ICs (representing 600 schools from 454 institutions), The \$29.19 median hourly wage estimate is derived from the Bureau of Labor Statistics’ “May 2019 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 (\$18.07 median hourly wage, representing 60% of ICs) and Education Administrators, Postsecondary (\$45.87 median hourly wage, representing 40% of ICs) figures.

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

Year	Description	Responses	Average Response Time	Total Burden Hours	Estimated Hourly Wage	Total Cost
2022	Doctorate recipients	52,440	20 minutes	17,480	\$41.00	\$716,680
	Institutional Contacts	600	20 hours	12,000	\$29.19	\$350,280
2023	Doctorate recipients	53,360	20 minutes	17,787	\$41.00	\$729,267
	Institutional Contacts	600	20 hours	12,000	\$29.19	\$350,280
Total Burden in 3 Years		107,000		59,267		\$2,146,507
Total Annual Burden		35,667		19,756		\$715,502

NCSES estimates that the average annual burden for the 2022 and 2023 survey cycles over the course of the three-year OMB clearance period will be no more than 19,756 hours at a cost of \$715,502.

A.13. Costs 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. Costs to the Federal Government

The cost to the Federal Government for this annual data collection is approximately \$1.9 million per year. This amount is based on the contract cost for the 2020-2021 SED survey cycles.

A.15. Changes from the Prior Cycle

The only known change in the 2022 and 2023 SED is an expected increase in SED universe size, resulting in a greater number of respondents. The new COVID-19 impact questions added in the 2021 SED increased the burden hours slightly, but it was offset by the revised education history section taking less time than the education section used in previous survey cycles.

A.16. Plans for Tabulation and Publication

The results of the SED will be disseminated in a number of ways. To release the data, NCSES will publish an annual report “Doctorate Recipients from U.S Universities” which includes a set of 72 data tables and a Digest summary report with approximately 36 figures highlighting findings from key survey themes. The data 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. 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 NCSES 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/builder/sed>). As stated previously, NCSES is developing a Restricted Data Analysis System that will allow users greater access to the SED data without compromising the confidentiality of respondents. This system, based on the NCES DataLab, will make SED data available beginning with data from the 2018 and 2019 SED. 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 2022 SED includes doctoral graduates from 1 July 2021 to 30 June 2022. The 2022 Web instrument will be uploaded upon OMB approval for continuous distribution to doctoral students as they complete their doctorate requirements. After the survey close-out, data variables will be constructed, edited, evaluated, and reviewed for trend consistency in January 2023. In February 2023, the file will be further evaluated, and quality control checks will be made. Data will be tabulated in April 2023 and prepared for publication by November 2023. Aggregate data will be made available to the public in December 2023 on the NCSES website.

Project Schedule

The 2022 SED survey schedule follows. The 2023 SED survey schedule is expected to be identical, except lagging by one year.

Phase	Time
Receive OMB clearance approval	April 2021
Survey instrument available to doctoral students	May 2021
Data collection close-out	December 2022
Preparation of data file	February 2023
Production of publications	April 2023
Release of data by NCSES	December 2023

A.17. Exception to Displaying of OMB Expiration Date

Not applicable. The OMB Control Number and expiration date will be displayed.

A.18. Exception to the Certification Statement

Not applicable. No exceptions to the certification statement are being sought.