

**SF-83 SUPPORTING STATEMENT**

**for**

**Survey of Earned Doctorates**

**for SED 2016 and 2017 survey cycles**

## TABLE OF CONTENTS

<b>SECTION A: JUSTIFICATION</b> .....	<b>4</b>
A.1. Necessity for Information Collection.....	4
A.2. Uses of Information .....	5
Current Uses of the SED at the Federal Level.....	6
Academic Uses of the SED.....	10
A.3. Consideration of Using Improved Technology.....	10
A.4. Efforts to Identify Duplication.....	12
A.5. Efforts to Minimize Burden on Small Business .....	14
A.6. Consequences of Less Frequent Data Collection.....	14
A.7. Special Circumstances .....	15
A.8. Federal Register Announcement and Consultations Outside the Agency .....	15
Consultations Outside the Agency.....	15
A.9. Payment or Gifts to Respondents.....	16
A.10. Assurance of Confidentiality .....	16
A.11. Justification for Sensitive Questions.....	17
A.12. Estimate of Respondent Burden.....	17
A.13. Cost Burden to Respondents.....	19
A.14. Cost to the Federal Government .....	19
A.15. Program Changes or Adjustments.....	19
A.16. Tabulation and Publication Plans and Project Schedule.....	20
Project Schedule.....	20
A.17. Display of OMB Expiration Date .....	21
A.18. Exception to the Certification Statement .....	21

<b>SECTION B: Collection of Information Employing Statistical Methods .....</b>	<b>22</b>
Survey Data Collection Procedures Background.....	22
B.1. Universe and Sampling Procedures.....	22
B.2. Survey Methodology.....	24
B.3. Methods to Maximize Response .....	24
B.4. Testing of Procedures.....	25
Data Collection Related Tests.....	25
Survey Quality Tests and Research .....	27
Proposed Tests and Research.....	30
B.5. Individuals Consulted.....	31

## **LIST OF ATTACHMENTS**

Attachment 1:	Proposed 2016 SED Questionnaire & Current 2015 Questionnaire
Attachment 2:	Changes to the SED 2016-2017 Questionnaire Compared with SED 2015
Attachment 3:	Authorizing Legislation of Sponsoring Agencies
Attachment 4:	Current Representatives from Sponsoring Agencies
Attachment 5:	List of Persons who reviewed the SED
Attachment 6:	Federal Register Announcement
Attachment 7:	NSF Staff and Contractor Data Use Agreement for Individuals
Attachment 8:	Materials Sent to Institutional Contacts
8.1	Example of the Institutional Profile
8.2	Example of the Institution Contact Administrative Tool (ICAT) (Web Collection Pin/Password for Institutions)
8.3	Institutional Contact Transmittal Form
8.4	Example of the Interim Results Form
8.5	Address Roster Form
8.6	Missing Information Roster
8.7	Dean/Contact Change Form
Attachment 9:	Example of Non-Respondent Follow-Up Letter
Attachment 10:	SED Methodological Research Listing

## SECTION A: JUSTIFICATION

This request is for OMB clearance for two years covering the 2016 and 2017 cycles of the Survey of Earned Doctorates (SED). The request represents an extension of a currently approved data collection (OMB No. 3145-0019). The SED underwent a review process this year which generated recommendations for a small number of changes to the questionnaire. The purpose of the review was to identify revisions and additions that would reduce respondent burden, improve data quality, increase consistency across agency surveys, reduce costs, and increase the efficiency of administrative processes in order to hasten the release of the survey data. The recommended changes from this review process and the reports supporting these recommendations are described in section B.4 of this report. Both the current (SED 2015) and recommended (SED 2016) questionnaire instruments are presented in Attachment 1. A list of methodological reports that have informed changes to the SED over the past 15 years appears in Attachment 10.

### *A.1. Necessity for Information Collection*

The SED began in 1957–58 to collect data annually on the number and characteristics of individuals receiving research doctoral degrees from accredited U.S. institutions. All individuals receiving research doctorates from accredited U.S. institutions 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. Doctorate recipients of professional doctorate degrees such as MD, DDS, JD, DPharm, and PsyD are not included in the survey. The results of this annual survey are used to assess characteristics and trends in doctorate education and degrees. This information is vital for educational and labor force planners within the federal government and in academia.

The SED is sponsored by the National Science Foundation in collaboration with the National Institutes of Health (NIH), the U.S. Department of Education (ED), the U.S. Department of Agriculture (USDA), the National Endowment for the Humanities (NEH), and the National Aeronautics and Space Administration (NASA). Sponsoring agencies typically provide funding for the SED, obtain special tabulations from the survey each year customized to their unique needs, and receive uniform data tabulations/reports. The representatives of each sponsoring agency and the list of persons who have been consulted and/or have reviewed the SED 2016-2017 questionnaire are listed in Attachment 5. The National Center for Science and Engineering Statistics<sup>1</sup> (NCSES) within the National Science Foundation has monitoring responsibility for the project, which is currently conducted under contract by NORC at the University of Chicago.

---

<sup>1</sup> NCSES is a federal statistical agency that reports to the National Science Foundation's Directorate for Social, Behavioral, and Economic Sciences. Formerly known as the Division of Science Resources Statistics (SRS), NCSES was established by the America COMPETES Reauthorization Act of 2010. The new name reflects broadened responsibilities for the collection, interpretation, analysis, and dissemination of objective data on the science and engineering enterprise.

NORC at the University of Chicago was competitively awarded a new procurement in 2012 that covers the SED operations from 2013 to 2016.

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, the NCSSES 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 R&D.

Statutory authority for collection of information for fields other than science and engineering comes from legislation for the other Federal sponsoring agencies. 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;
4. USDA: Title XIV of the Agriculture and Food Act of 1977 (PL 95-113) as amended, and Title V of the Equity in Educational Land-Grant Status Act of 1994 (PL 103-382) as amended, Sec. 354;
5. NASA: Title 42 of The Public Health and Welfare and Chapter 26 of the National Space Program.

Attachment 3 provides the cited legislation for each sponsoring agency; these agencies may change depending on their continuing funding for the SED.

## ***A.2. Uses of Information***

The SED has been conducted annually since 1958 and is jointly sponsored by six federal agencies in order to avoid duplication in data collection. The purpose of the SED is to compile data on all recipients of earned research doctorates awarded by U.S. universities. It is an accurate, timely source of information on one of our nation's most precious resources – highly

educated individuals. The burden on respondents is relatively light and the resulting information is used extensively and further analyzed by many federal agencies for program evaluation and policy formulation.

The SED is the only data source that provides comprehensive information on the education and early career commitments of persons who have recently received research doctorates from U.S. universities. The resulting information is a valuable resource for other government agencies, academic researchers and policymakers, as well.

The results of the SED each academic year become part of the Doctorate Records File (DRF), a virtually complete database of more than 1.9 million U.S.-educated doctorate recipients from 1920 to 2013.

The six sponsoring agencies have made extensive use of the SED. Detailed tables, tabulations, and the computer files are available to representatives of the agencies that sponsor the SED for use in program planning/evaluation, policy development, and dissemination. The heads of the agencies use the data in their reports and speeches, as well as in national forum discussions of educational policy.

There is no public-use SED file available; however, organizations and individuals can request information from the SED database and these special tabulations are provided at cost by the survey contractor. In addition, selected SED data items are available to the public on the internet through the WebCASPAR database (<http://webcaspar.nsf.gov>) and the SED Tabulation Engine (<https://nces.norc.org/NSFTabEngine/#WELCOME>). Doctorate-granting institutions participating in the SED may receive a cumulative micro-level data file of their own institution's doctorate recipients free upon request. Doctorate-granting institutions also receive preliminary institution files during the course of a survey cycle, also free upon request. Researchers at U.S. institutions may gain access to the cumulative micro-level data file by completing an NCSES Restricted-Use Data Licensing Agreement (<http://www.nsf.gov/statistics/license/>). Statistical data from the SED are widely used by other federal agencies, Congress, state agencies, universities, professional societies, and individuals doing research in science policy, graduate education, economics, and human resource planning.

### ***Current Uses of the SED at the Federal Level***

The uses made of the data from the SED reflect the fact that the survey is the most comprehensive, accurate, and timely source of data on research doctorate degree awardees in the United States. The use of SED data and reports is widespread among sponsoring federal agencies and other federal organizations. The data are used for policy development, in carrying out program responsibilities of the agencies, and in the administration of agency programs. The data are disseminated extensively throughout the agencies. Some of the more important recent uses, organized by user agency, are listed below. The participating federal agencies are subject to change, pending funding availability; the current liaisons for each sponsoring agency are listed in Attachment 4.

#### ***a. The National Science Foundation***

The National Science Foundation has been a sponsor of the SED since 1958. The uses made of the data on science and engineering (S&E) doctorates are many:

- The SED is used as the universe frame for selecting the sample of doctoral scientists to be included in the NSF's Survey of Doctorate Recipients (SDR), a longitudinal survey of doctorate recipients in science, engineering, and health fields.
- Many programs within the NSF, especially those dealing with women, minorities, and persons with disabilities, use data from the SED for program planning. While these programs focus on U.S. citizens, data on foreign citizens studying here for their PhD are also useful for international comparisons and for quantifying the attraction of the U.S. graduate education around the world. In addition, the SED has served as a measure of program effectiveness, in that the Graduate Research Fellowship Program has used the information on those who complete the PhD to evaluate the effectiveness of the program and its design requirements.
- Several reports are published on science and engineering doctorates by the NSF for internal and external use. The first report to be released each year is available publicly in November, eleven months after survey closeout. Additional reports follow which provide more detailed data or more analysis of the results from the SED.

Special tabulations of data from the survey constitute a key resource in meeting policy and program information needs of the Foundation. Examples of uses within the Foundation include:

- Data on doctorates awarded to minorities and women for presentation to the National Science Board for their use in examining participation of these groups.
- Data on foreign scholars provided to an interagency committee studying foreign access to U.S. science and engineering at American colleges and universities.
- Data in support of the Careers of Doctorate Holders project, an international effort led by the Organization for Economic Co-operation and Development (OECD) to improve the quality and comparability of international data on doctorate holders.
- Baccalaureate institutions of science and engineering doctorate recipients supplied to the NSF's Division of Undergraduate Education for use in a study of institutions' contributions to the highly trained labor force.
- Data supplied by the NSF to outside users. At the national level, within recent years, major data users have included the White House Office of Science and Technology Policy, the National Academy of Sciences, and others.
- Published results in widely distributed NSF publications. Data are included in two of the Foundation's Congressionally-mandated biennial reports, *Science and Engineering Indicators* (<http://www.nsf.gov/statistics/seind14/>), and *Women, Minorities and Persons*

*with Disabilities in Science and Engineering*  
(<http://www.nsf.gov/statistics/wmpd/2013/start.cfm>).

- A range of topics related to doctorate recipients addressed in selected data tables that are available electronically on the NCSES Web Site ([www.nsf.gov/statistics/doctorates](http://www.nsf.gov/statistics/doctorates)).

Aggregated data on selected variables are available for each institution through the WebCASPAR database, available to the public (<http://webcaspar.nsf.gov>). A complementary tool, the SED Tabulation Engine, provides additional variables to the public at (<https://nces.norc.org/NSFTabEngine/#WELCOME>).

### ***b. The National Institutes of Health***

For more than 30 years, the National Institutes of Health (NIH) has used the results of the SED to meet a variety of planning, evaluation, 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 track 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.** By allowing comparisons and sustained tracking of selected doctoral recipients, the SED is a critical tool in Government Performance and Results Act (GPRA) reporting on the effectiveness of NIH predoctoral training grants. In addition, since 2008, NIH has used the results of the SED 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***

The U.S. Department of Education (ED) has sponsored the SED since 1958. The Department's National Center for Education Statistics (NCES), Postsecondary Education Statistics Division, funds 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. Trend data are compiled each year and displayed in tables in the Center's publication *Digest of Education Statistics*. NCES has also published a report containing tables from the Doctorate Records File comparing education doctorates to doctorates in other fields, by selected characteristics.

Data from the SED are also used for evaluation by ED's programs, such as the Office of Student Financial Aid, the individual program offices, and by the Office of Planning, Evaluation, and Policy Development.

#### ***d. The Department of Agriculture***

The U. S. Department of Agriculture (USDA), a sponsor since 1988, has developed a list of discipline areas in which it has particular concerns, analogous to the subsets developed by the National Science Foundation and the other sponsors, and has requested trend tabulations on doctorate recipients in these fields. Data collected in the SED are used to evaluate how widespread these programs and fields are in the United States. Data are also used in the evaluation and planning of 1890 Land Grant and Tribal College programs. A considerable and expanding number of tabulations from the SED are also available on their Food and Agricultural Education Information System (FAEIS) and other websites.

#### ***e. The National Endowment for the Humanities***

The authorizing legislation for the National Endowment for the Humanities (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 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.

#### ***f. The National Aeronautics and Space Administration***

As a sponsor since 1995 of the SED, the Education Division of the National Aeronautics and Space Administration (NASA) has developed a program for the utilization of data from the SED in its planning and information dissemination activities. SED data have been useful in providing information on the progress of women and minorities in science and engineering.

#### ***g. Other Federal Agencies and Congress***

Other Federal agencies have utilized the SED in several ways – through requests for special tabulations and tables, data requests from NSF, and licensing agreements. Congressional staffers have contacted NSF for information on 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 Uses of the SED***

The nation's doctorate-granting institutions not only contribute to the SED data set but also utilize the data for many purposes. Each year since 1997, the sponsors have provided to the dean of each graduate school profiles of their doctorate recipients' demographic characteristics, debt status, postgraduation plans, employment and other activities, compared with national and peer-institution data (see Attachment 8.1 for an example of an Institutional Profile).

Graduate and baccalaureate institutions use the data in program planning, comparison with other institutions or with national figures, and in the development of affirmative action plans. SED data on the number of research doctorates awarded to racial/ethnic minorities are used extensively by institutions as the only reliable source of the supply of persons with particular characteristics for academic positions. Site visits have confirmed the usefulness of the data to institutions. In the past few years, NSF has seen the increasing use of data by graduate deans to address issues of changes in the composition of the graduate schools and the time it takes to complete the doctorate, etc., and they have relied on the SED as the "ready-made" database on their graduates.

Over the past two years, almost 120 requests for data by graduate deans, other academic administrative offices, and individual researchers have been fulfilled, 61 of which were requests for institutional datasets. In 2009 NSF eliminated the fee for these datasets, allowing institutions even greater access to their own data. The SED also provides preliminary institution files to graduate deans during the course of a survey cycle, allowing institution researchers more immediate access to their graduates' data for analysis. The U.S. universities help administer the SED to their doctorate recipients, and it is clear that they get something back from the survey because they can use the data for their own purposes; this is a symbiotic data collection effort.

In addition, researchers can apply to use selected microdata from the SED under the NCSES Restricted-Use Data Licensing Agreement if publicly available data do not address the specific needs of their study. The NCSES Licensing Agreement is executed between an institution and NCSES. The NCSES license is a legal document that requires that stringent security procedures be taken to ensure that the confidential data will be protected against unlawful disclosure.

#### ***A.3. Consideration of Using Improved Technology***

Planning for the development of an electronic questionnaire administration of the SED was begun in late 1999 and has been refined, implemented, and expanded since that time. The purposes of instituting an electronic, web-based option were to ease the burden on students by offering alternative completion modes, to help assure continued high response rates, and provide an option for institutions using electronic graduation packages.

The population for this survey, graduating doctoral students, is virtually all computer literate and familiar with the internet. Offering a paperless survey version which can be submitted electronically is not only appealing to many students, it is also very practical for respondents

who are often relocating at the close of their studies and are not near the graduate offices for submission of their completed questionnaires.

The phase-in of the SED electronic procedures was gradual and consisted of three parts: a web survey instrument to parallel the paper version; non-respondent follow-up via email for missing questionnaires; and a web interface for institutions, which is password-protected for each school (see Attachment 8.2). The web version of the instrument was carefully developed and tested to assure that mode effects between it and the paper version were minimal and non-biasing. Students access the web survey directly upon registering with the SED and are also provided with a PIN/Password via email should the student decide to complete the survey at a later date. Students who do not complete the survey before graduation receive a PIN/Password via an email or letter prompt. Because web survey design has evolved since 1999, an expert review of the web-based data collection instrument occurred in 2009, with suggested changes implemented in the 2010 SED survey cycle. Additionally, in 2011, a set of NSF-supported methodologists reviewed the SED web-instrument as part of a comprehensive review of NSF web surveys. Most of the recommendations from this review were incorporated in the 2014-2015 questionnaires. Finally, further methodology work is continuing in advance of the 2016 data collection effort to modernize the design of the web instrument, including updating its general appearance, advancing functional operations, and establishing flexibility across web platforms (including mobile and tablet).

NSF has made a Web Institution Interface (WII) available to all participating institutions that allows Institutional Contacts (ICs) to enter their own password-protected site to monitor completion rates for their graduates, link to various SED reports, and print questionnaires from PDF document files. For the 2014 survey cycle, the WII was redesigned to improve the interface and enhance usability and was renamed the Institutional Contact Administrative Tool (ICAT). The ICAT also allows ICs to compare their list of graduates, and their completion status, with the SED contractor-maintained database in order to track the response rate for their institution. The ICAT includes two levels of interaction. Level 1 includes general information and communication, including a calendar with important survey dates, a forms library with generic and publicly available forms, and a multimedia library with instructional videos. Level 2, which is institution specific and only available to each unique IC user, includes current student roster data, address rosters, and missing information rosters. ICs review these data and upload revised files with student-level updates as well as review the current survey completion status of active rounds.

The web-based data collection system is becoming more widely accepted by both graduates and institutions. In the 2013 SED, 84 percent of completed questionnaires were completed in the web mode. This demonstrates a steady increase from 2011 and 2012, when web questionnaires constituted 39 percent and 63 percent, respectively, of all completed questionnaires. The web completes include students who graduate from schools that use the web survey as the primary or secondary method of completion, and students who are reached via the non-respondent follow-up process. NSF expects that the proportion of students using the web version of the SED will continue to grow as more graduate schools move to web-based systems for managing the degree completion paperwork of their doctoral students.

The degree-awarding institution is the main SED interface with the doctorate recipient. The IC at the institution (usually located in the graduate dean's office) helps administer the survey, either via the registration URL or the hard copy, tracks the completion status, collects and submits the completed questionnaires and final graduation lists to the SED survey contractor. However, because the SED work is just one of many duties performed by the ICs, SED data collection staff adjust the IC coordination tasks, as needed, to fit with each institution's procedures for processing and awarding doctoral degrees. The work of maintaining the overall cooperation of the schools while phasing in web-based systems remains a continuing challenge. However, the final decision on how best to conduct the SED is left to the administrators at each participating institution based on their individual preferences and capabilities.

Over the past few years, the SED has made significant efforts to increase web participation among the largest SED schools. As part of this outreach, a number of schools were offered direct assistance from SED staff, who explained the web process, answered questions, and provided technical support. As an added incentive, these schools were also given the opportunity to receive, shortly after the graduation period, the unedited data from their institution's respondents who had completed the web version of the SED during the academic term. The quick availability of this unedited SED data enables institutional researchers at participating universities to analyze SED data contemporaneously with data collected by their own surveys of students (e.g., graduate student "exit surveys"). In the normal SED data processing cycle, the fully edited SED data on an institution's recently graduated doctorate recipients, complete with the data collected via the survey contractor's follow-up procedures, is made available to universities 17 months after the end of the academic year. A small but growing number of institutions are taking advantage of the opportunity to access and analyze unedited SED data soon after the end of the graduation period. To further facilitate transition to the web survey mode, institutions are also able to link the SED web survey to their institution-specific exit survey, which enables students to move immediately from the completion of one survey into the initial web-page of the other survey. This feature not only reduces respondent burden, it will reduce the need of both the survey contractor and the institutions to prompt students to complete the surveys.

#### ***A.4. Efforts to Identify Duplication***

During collaborations with other agencies, the National Science Foundation has reviewed and determined that no other government survey gathers identical or even similar information. In addition, the National Science Foundation actively maintains contacts with professional societies and groups, such as the Council of Graduate Schools, within both the higher education and data collection communities, so that information about any surveys similar to the SED would be immediately known.

SED survey content is coordinated with NSF's Survey of Doctorate Recipients (SDR) and with the SESTAT data collections on scientists and engineers. The SDR is designed to provide demographic and career history information about a sample of individuals with doctoral degrees in science, engineering and health fields. The results of the SDR are vital for educational planners within the federal government and in academia. The SDR results are also used by

employers in all sectors (education, industry, and the government) to understand trends in employment and salaries in S&E fields for doctorate holders and to evaluate the effectiveness of equal opportunity efforts. NSF finds the results important for internal planning because most NSF grants and fellowships go to individuals with doctoral degrees. The Doctorate Records File from the SED is used to identify SDR respondents in science, engineering and health fields. Contact information obtained in the SED is necessary for contacting new PhD's, who are added to the SDR sample every two years. The coordination of content and procedures is, therefore, critical to the success of both the SED and SDR surveys.

The database system known as the Scientists and Engineers Statistical Data System (SESTAT) combines data from the SDR, the National Survey of College Graduates, and the National Survey of Recent College Graduates (now defunct). The SESTAT system is designed to provide a comprehensive picture of the number and characteristics of individuals with training and/or employment in science and engineering in the United States. The SED survey content is coordinated with the SESTAT surveys to avoid unnecessary duplication of items and to assure relevant uniform approaches on similar items such as race/ethnicity and specific functional limitations.

Differences between the SED and the Integrated Postsecondary Education Data System (IPEDS), which collects some information on doctoral degrees, are outlined below. 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 gender. IPEDS provides data on aggregates of institutional doctorate recipients (including race/ethnicity and gender) and not data by individuals.

The SED obtains information from the individual research doctorate recipient on over two dozen variables, information not collected through the IPEDS survey. As mentioned earlier, NCES uses the SED to present data that are not available from IPEDS.

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

- In the SED, field of degree, citizenship, gender, and race/ethnicity are frequently used in analyses that link these variables with other key variables, such as the length of time spent pursuing the degree and the amount of debt accumulated during graduate education. These other linked variables cannot be collected from the institutions that provide information to IPEDS. The field of degree and demographic variables are also used to identify individuals in "rare subgroups" for oversampling in the SDR (described above). Without these questions, the SDR would need to be greatly expanded to meet the needs for the congressionally mandated report, *Women, Minorities and Persons with Disabilities in Science and Engineering*, for education and labor market data.
- It is not a feasible option to exclude collection of the information about doctoral degree recipients from IPEDS, because inclusion of field of degree, citizenship, gender, and

race/ethnicity permits comparative analyses of trends in degree production at different degree levels. SED data cannot be substituted for the IPEDS in such comparisons because of the inevitable differences between an institutional survey and a demographic survey. For example, individuals' racial/ethnic self-identification on these variables may differ from those maintained by the institutions. Also, as IPEDS collects data on types of doctoral degrees that are explicitly excluded from the SED – the “doctorate degrees-professional practice” and the “first professional” types of doctoral degrees – eliminating doctoral degrees from the IPEDS data collection would result in a loss of information about these important types of degrees.

- Including field of degree, citizenship, gender, and race/ethnicity questions on both surveys also 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 academic fields of study and participation in disciplines by demographic groups of interest (including U.S. and non-U.S. citizens on both permanent and temporary visas). These data 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 administration of the survey in the graduate deans' offices. The survey collects data from each person receiving a research doctorate at the time they complete the requirements for their degree. Staff at the graduate deans' offices include the link to the web survey with their other electronic documents, post the PDF version of the survey on their website, 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 of the survey collection process is imperative for the usefulness of the data to the federal agencies and for the ease of collection by universities. A continuation of the current survey methodology serves the best interests of all involved.

If the SED were conducted less frequently, there would also be significant repercussions to the success of the Survey of Doctorate Recipients (SDR). The Doctorate Records File is the sample frame used to identify SDR sample members. Locating information obtained from the SED is necessary for contacting the new PhDs who are added to the SDR sample. The coordination of

timing, content, and procedures of these two studies is, therefore, critical to the success of both the SED and SDR surveys.

#### ***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 November 24, 2014 (see Attachment 6). No public comments were received during the 60-day comment period.

#### ***Consultations outside the Agency***

In the many years of operation of the SED, the six federal sponsors and the survey contractors have consistently invited others to comment on the SED. The comments come from the federal sponsors themselves, an expert panel convened by NCSES, the Council of Graduate Schools, and other governmental and academic institutions. Comments and suggestions regarding the SED and the manner in which it is conducted have been received from individual respondents, university faculty advisors, graduate deans' offices, and professional researchers by telephone, email, mail, and in-person contacts. University representatives have been sought out for consultation at venues such as professional conferences, meetings and personal site visits to institutions. These consultations have helped to determine if there are problems in the conduct of the survey or in the interpretation of certain items. These problems are discussed with graduate deans for their conceptual validity and applicability to all fields of study, and the need for such information is weighed against respondent burden.

The federal sponsors meet together at least twice a year to discuss the SED design, operation and dissemination activities, and to plan future activities. The federal sponsors also 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 decision made every two years on taxonomy changes for the SED.

NCSES has convened multiple meetings of a Human Resources Experts Panel (HREP) in order to help improve data collection on the education and employment of the S&E labor force through review and renewal of the S&E personnel 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 both researchers and policy analysts 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; 6) identifying ways in which S&E personnel data could be more useful and relevant for analyses; and 7) proposing ways to enhance

the content of the NCSES human resources 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. The SED was the main topic at the January 2010 meeting which was focused on eliciting the expert panelists' ideas for prospective questionnaire items to be included in future versions of the SED.

Formal site visits have been conducted by National Science Foundation and survey contractor staff for the purpose of consulting with graduate deans and campus administrators. The institutions visited include those with poor response rates, primarily to resolve the survey collection problems at those institutions. However, the site visits also allow for the discussion of the uses of the SED by the federal sponsors and by the universities themselves.

### Other Consultations

The SED has also been informed by numerous other contacts between NSF and 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, and the Organization for Economic Co-operation and Development. Routine information requests also provide insight into the interests of the general public.

NSF has also held several meetings with representatives of the American Association of Universities (AAU), a nonprofit association of 60 U.S. and two Canadian research universities, to discuss the feasibility and desirability of linking the web-SED instrument to the web-based exit-survey instruments of AAU institutions, merging SED data with exit-survey data, and incorporating commonly-used exit-survey items within the SED.

### ***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 uses to be made of the data and the voluntary nature of his/her responses. The confidentiality pledge to the SED respondents is:

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. 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, 4201 Wilson Blvd., Arlington, VA 22230, 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, and all project staff are required to sign confidentiality agreements before they first have access to any SED data, and on a yearly basis thereafter (see Attachment 7). Data files with personal identifiers are provided to NIH, a federal sponsor, and its contractors. NIH receives the personal identifiers only after they have executed an NCSES Restricted-Use Data Licensing agreement and all contractors have signed 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. No one outside of these groups can obtain data files with direct identifiers such as phone numbers and addresses. Qualifying researchers can obtain SED microdata (but no direct identifiers) only by executing an NCSES Restricted-Use Data Licensing Agreement with NSF through their employer.

#### ***A.11. Justification for Sensitive Questions***

The SED recognizes the growing sensitivity of requesting respondents' Social Security numbers to an increasing segment of the population. The SED is allowed 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 only collects the last four digits of the Social Security number to be used to ascertain the correct identity of the survey respondent in survey operation and evaluation purposes.

#### ***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 2016, approximately 55,000 individuals are expected to receive research doctorates from U.S. institutions. The estimated average response time for the 2016 SED, 20 minutes, was based on cognitive interviews and tests conducted by the survey contractor. Given

the lack of substantial changes between the 2016 SED and 2015 SED – one follow-up question and one response option added and the wording of two response options modified – the estimated average response time for the 2016 SED remains 20 minutes. An expected response rate of 92% would indicate 50,600 respondents. Thus, the total respondent burden for completing the survey questionnaire is estimated to be 16,867 hours.

The estimated cost to respondents for this data collection is \$486,613, based on the estimated 16,867 hours of response burden at a time-cost of \$28.85 per hour. The \$28.85 per hour time-cost estimate is derived from results from the 2013 SED which indicate that the median income for doctorate recipients was \$60,000. Assuming a 40-hour work week and 52 weeks of work per year, this \$60,000 income per year figure works out to \$28.85 of income per hour.

In addition to the actual survey, the SED also requires the collection of administrative data from participating institutions. The IC at the institution helps distribute the registration URL or hard copy survey, track it, collect it, and submit the completed questionnaires and graduation lists to the SED survey contractor. This requires that the following materials be sent to the ICs (see Attachments 8.3 through 8.7 for examples of the materials):

- Transmittal Form, to be included in the package of completed surveys or with a graduation list, asks for the total number of graduates for the graduation date in question, the total number of surveys enclosed, and contact information for each non-respondent;
- Interim Result Form, reports the number of graduates currently accounted for on each graduation date;
- Address Roster Form asks for a physical mailing and/or e-mail address for non-respondents that were not already provided on the Transmittal Form;
- Missing Information Roster asks for the critical items for any non-respondents and the missing critical items for respondents;
- Dean/Contact change form is available on the ICAT for the institution to submit any change to the contact information of the dean or IC.

Based on focus groups conducted with ICs, it is estimated that the SED demands no more than 1% of the IC’s time over the course of a year, which computes to 20 hours per year (40 hours per week x 50 weeks per year x .01). At an estimated average wage rate of \$29/hour for ICs, and with 570 participating programs, the total estimated time-cost to ICs of administering the SED is \$330,600 per year. The \$29/hour wage estimate is derived from the Bureau of Labor Statistics “Occupational Employment and Wages, May 2013” for a combination of Office and Administrative Support Occupations (60% of ICs) and Education Administrators, Postsecondary (40% of ICs).

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	Annual Cost Burden
Doctorate Recipient SED Questionnaire	50,600	20 minutes	16,867	\$486,613

Institutional Contacts	570	20 hours	11,400	\$330,600
<b>Total SED Annual Burden</b>			<b>28,267</b>	<b>\$817,213</b>

NSF anticipates conducting methodological research over the clearance cycle that would involve respondents (see Section B.4 for description of possible methodological research). These tasks would most likely involve cognitive interviews of as many as 100 individuals. NSF plans to conduct this research under the Generic Clearance of Survey Improvement Projects (OMB Control #3145-0174).

***A.13. Cost Burden to Respondents***

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

***A.14. Cost to the Federal Government***

The cost to the Federal Government for this annual data collection is approximately \$3.76 million per year. This amount was based on the negotiated contract cost for the 2016 SED.

***A.15. Program Changes or Adjustments***

There are two adjustments in the SED 2016 that will impact data collection processes and costs. The first is the annual expected increase in SED universe size, which would result in an increase in survey costs due to the higher number of respondents. The second is the removal of the Missing Information Letter (MIL) as a survey data collection mode, which will result in a slight decrease of per respondent burden. In regards to institutions, there are no adjustments to the data collection processes that would impact the per institution burden indicated in A.12 above.

***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, NSF 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 status, sex and disability status. NSF will also publish a Summary Report Digest, a short publication with approximately 30 figures highlighting many of the important findings from the survey. The Digest will be available in both print and electronic formats.

The printed Digest is provided free of charge to responding institutions participating in the SED 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 become available on NSF's NCSES website.

Aggregated data on selected SED variables are publicly available through the WebCASPAR database (<http://webcaspar.nsf.gov>). SED data for race/ethnicity, sex, and citizenship for 2007 and later years are no longer available in WebCASPAR but are available in the SED Tabulation Engine, a tool developed in early 2011 that includes a disclosure control mechanism that protects the identity of respondents when reporting gender, race/ethnicity and citizenship variables. Additionally, SED data will be available to researchers via the NSF's data enclave, a secure environment (currently in the pilot stage) that provides researchers remote access to microdata while still protecting respondent confidentiality.

The Scientists and Engineers Statistical Data System (SESTAT), which resides on the Web, includes some variables taken from the SED that pertain to SDR respondents (e.g., past educational degrees). The SESTAT system, described in section A.4, can be used to produce tabulations from the component surveys, providing a rich resource to those within and outside the government. As noted above, microdata are also disseminated to federal co-sponsors and collaborating researchers (with current data use licenses) so that they can conduct specialized studies. These, in turn, are expected to result in reports and other publications that further disseminate the data. Finally, it is anticipated that substantive analyses of the SED data will be presented at appropriate 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, etc.

The SED project schedule was shortened starting with the 2013 round. The 2016 survey covers the period from July 1, 2015 to June 30, 2016. The paper SED questionnaires 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 the requirements for their doctorate. Returned survey questionnaires will be edited and coded until survey close-out, which is December 2016 for the 2016 academic year (SED 2016). After the survey close-out, data variables will be constructed, edited, evaluated, and reviewed for trend consistency in January 2017. In February 2017, the file will be further evaluated and quality control checks will be made. Data will be tabulated in April 2017 and prepared for publication by November 2017. Aggregate data will be made available to the public in November 2017 via the on-line Summary Report Data Tables on the NCSES website.

### ***Project Schedule***

The 2016 SED survey schedule follows. The 2017 SED survey schedule is expected to be similar except lagging by one year.

**Phase**

**Time**

Receive OMB clearance approval	May 2015
Survey Instrument available to students	May 2015
Data collection close-out	December 2016
Preparation of data file	February 2017
Production of tabulations	April 2017
Release of data by NSF	November 2017
Interagency Report released	November 2017

***A.17. Display of OMB Expiration Date***

The OMB Expiration Date will be displayed, as indicated.

***A.18. Exception to the Certification Statement***

The 2016-2017 SEDs will comply with the certification statement on form OMB 83-1.