

SF-83-1 SUPPORTING STATEMENT

For

2014 Pilot Early Career Doctorates Survey

Section	Page
A. Justification.....	1
A.1 Need for Data Collection and Legislative Authorization.....	3
A.2 How, by Whom, and for What Purpose the Information Is to Be Used.....	3
A.3 Consideration of Using Improved Technology.....	3
A.4 Identification of Duplication.....	4
A.5 Small Businesses Involvement.....	5
A.6 Consequences of Less Frequent Surveying.....	5
A.7 Special Circumstances.....	5
A.8 Federal Register Notice and Consultations outside the Agency.....	5
A.9 Payment or Gifts to Respondents.....	5
A.10 Assurance of Confidentiality.....	6
A.11 Sensitive Questions.....	7
A.12 Estimated Response Burden.....	7
A.13 Cost to Respondents.....	8
A.14 Cost to the Federal Government.....	8
A.15 Changes in Burden.....	8
A.16 Project Schedule for Information Collection and Publication.....	8
A.17 Displaying the OMB Expiration Date.....	9
A.18 Exceptions in Item 19 on Form 83-1.....	9
B. Collection of Information Employing Statistical Methods.....	10
B.1 Universe and Sampling Procedure.....	10
B.2 Description of Survey Methodology and Statistical Procedures.....	13
B.2.1 Imputation for Item Nonresponse in the Pilot ECD Survey.....	14
B.3 Methods Used To Maximize Response Rate.....	15
B.4 Testing of Procedures.....	16
B.4.1 Test of Building Sample Frame.....	16
B.4.2 Test of Contacting ECDs.....	17
B.4.3 Field Test of the Questionnaire.....	19
B.4.4 Changes to the ECD Questionnaire.....	19
B.5 Names and Telephone Numbers of Individuals Consulted.....	19

TABLES

Number		Page
Table 1.	Pilot ECD Survey Estimated Burden by Stage and Respondent Type.....	8
Table 2.	Pilot ECD Survey Data Collection and Publication Schedule.....	9
Table 3.	Sampling Strata and Estimated Sample Sizes for the 2013 Pilot ECD Survey.....	11
Table 4.	Institutions providing ECD Listings, Sample Members, and Response and Eligibility Rates Achieved in the Methodological Study by ECD Sampling Strata.....	12
Table 5.	Sample Sizes and Eligibility Rates Projected for the Pilot ECD Survey.....	13
Table 6.	E-mail Subject Line Experimental and Control Conditions.....	18
Table 7.	Individual Consulted on Pilot ECD Survey Technical and Statistical Issues.....	20

LIST OF ATTACHMENTS

- Attachment A. Contact Strategy Narrative
- Attachment B. Institutional Contact Materials
- Attachment C. ECD Contact Materials
- Attachment D. Confidentiality and Security Measures
- Attachment E. ECDS Questionnaire
- Attachment F. Federal Register Announcements

A. JUSTIFICATION

The National Center for Science and Engineering Statistics (NCSES) of the National Science Foundation (NSF) serves as a central Federal clearinghouse for the collection, interpretation, analysis, and dissemination of objective data on science, engineering, technology, and research and development for use by practitioners, researchers, policymakers, and the public. The Early Career Doctorates Survey will become part of an integrated survey system that meets the human resources part of this mission.

The Early Career Doctorates Project was established to gather in-depth information about early career doctorates (ECDs), including postdoctoral researchers (postdocs). Early career doctorates are critical to the success of the U.S. scientific enterprise and will influence U.S. and global scientific markets for years to come. Despite their importance, current surveys of this population are limited, and extant workforce studies are insufficient for all doctorates who contribute to the U.S. economy. The NSF's Survey of Earned Doctorates (SED) and the Survey of Doctorate Recipients (SDR) are limited to individuals who received their research doctorates from U.S. academic institutions, thereby excluding individuals who earned their professional doctorates and those who earned doctorates from institutions outside the United States but are currently employed in the United States. The NSF's Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS) provides aggregate level data for all postdocs and nonfaculty researchers regardless of where they earned the degree. However, the GSS is limited to science, engineering, and selected health (SEH) fields in U.S. academic institutions and their related research facilities and is collected at the program rather than individual level.

Through its multi-year Postdoc Data Project, NCSES determined the need for and the feasibility of gathering information about postdocs and other ECDs working in the United States.

Efforts to reliably identify and gather information about postdocs at the individual level have proved difficult due to substantial variation in how institutions characterize and report postdoc appointments. As a result, NCSES expanded the target population to include all individuals who earned their first doctoral degree within the past 10-years. Expanding the population to doctoral degree holders within the past 10 years ensures a larger, more consistent and reliable target population. Unique in scope, the key goals of the ECD Project are:

- To broaden the scope and depth of national statistics on the ECD population both U.S. degreed and non-U.S. degreed, across various employment sectors and fields of discipline,
- To collect nationally representative data from ECDs that can be used by funding agencies, policy makers, and other researchers to better understand the labor markets and work experiences of recent doctorate recipients, and
- To establish common definitions for different types of ECDs (e.g., postdocs, junior faculty, and other nonfaculty researchers) that can be applied across and within employment sectors.

The current focus of the Early Career Doctorates Project is to conduct a survey of ECDs working in three areas of employment: U.S. academic institutions, Federally Funded Research and Development Centers (FFRDCs), and the National Institutes of Health (NIH) Intramural Research Programs (IRP). NCSES, under generic clearance (OMB #3145-0174), conducted the

ECD study, a methodological study designed to test the feasibility of obtaining lists of ECDs from these three areas of employment and then sampling and surveying ECDs based on these institutional lists¹. The methodological study was designed to test a data collection strategy that uses institutional contacts as the conduit for questionnaire dissemination to ECDs working in the above employment settings. This data collection strategy will be used in the pilot survey of ECD (Pilot ECD Survey). The Pilot ECD Survey will be a two stage sample survey design.

Completed in February 2013, the methodological study provided substantial evidence that institutions in the target population can build sampling frames for their ECDs and that NCSES can successfully contact and survey these individuals through the information on these frames. As of the close of data collection:

- 74 of 81 institutions had agree to participate,
- 56 institutions had provided lists of their ECDs,
- 2,115 individuals had been sampled from these lists and asked to complete a 30-minute questionnaire, and
- 1,294 respondents had completed the survey.

In total, 71.8% of sampled individuals responded to the survey invitation in some way and 70.6% responded sufficiently to determine whether they were an ECD. Of the respondents in the 39 institutions that had highest degree and year of highest degree on their administrative databases, 97.8% were ECDs. Of the respondents from the 17 institutions that did not have highest degree and year of degree for all employees, 92.3% of those sampled from the likely ECD stratum were confirmed to be ECDs, 72.3% of those sampled from the possible ECD stratum were ECDs, and 18.8% of those sampled from the unlikely ECD stratum were ECDs.² The overall number of ECDs (or possible and likely ECDs) listed by participating institutions corresponded closely with the number ECDs estimated prior to the survey based on the GSS and IPEDS data. Taken together, these findings support the premise that participating institutions can provide high quality information about the ECDs working at their institution and that NCSES can use this information to sample and survey ECDs within the target population.

The Pilot ECD Survey will incorporate design changes based on findings from the methodological study to improve response rates, maximize survey coverage, and reduce burden per respondent. Through an increased sample of institutions, NSF will produce national estimates on ECDs working in U.S. academic institutions, FFRDCs, and the NIH IRP. Results from the Pilot ECD Survey will be used to fine tune the full implementation of the Early Career Doctorates Survey.

¹ NCSES is currently assessing the feasibility of building a frame of non-academic organizations employing ECDs. If successful, a methodological study may be conducted in the future to evaluate the feasibility of expanding the current frame to include these non-academic organizations in the data collection.

² Institutions that did not have highest degree and year of degree for all employees were asked to provide listing of employees for job titles that might include ECDs. Individuals in these lists were then coded into one of three categories – likely ECD, possible ECD, and unlikely ECD, based on the information provided, including highest degree earned, job title, and age.

A.1 Need for Data Collection and Legislative Authorization

The authority to collect information for the Pilot ECD Survey 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 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 R&D.

The Pilot ECD Survey will greatly enhance the center's current postdoc-related data collections and provide critical information that will be used to meet NSF's responsibilities under the Acts and the Executive Order.

A.2 How, by Whom, and for What Purpose the Information Is to Be Used

The proposed data collection will yield the first national estimates of the total ECD population employed by U.S. academic institutions, FFRDCs and NIH IRPs. In addition, the data will provide estimates of this population by gender, citizenship, origin of doctoral degree, and field of study. NSF will first publish statistics from the survey in the NCSES' InfoBriefs then in reports, including the NCSES' Science and Engineering Indicators and the Women, Minorities and Persons with Disabilities in Science and Engineering reports.

Restricted-use and public use data files will also be developed. Public-use data will be available through NSF website; access to restricted-use data files will be assessed and granted on a case-by-case basis with licensees signing affidavits of nondisclosure. Likely users of the data include the National Postdoctoral Association (NPA), the American Association for the Advancement of Science (AAAS), the Council of Graduate Schools (CGS), the American Association of Medical Colleges (AAMC), the Association of American Universities (AAU), the National Institutes of Health (NIH), and the Organization for Economic Cooperation and Development (OECD). Possible uses of the data include descriptive statistics on the employment and productivity of postdocs and other ECDs, the pathways followed by ECDs, and their future career plans, including plans to work outside the United States.

A.3 Consideration of Using Improved Technology

The primary form of communication and response for both institutional and individual participation is electronic. At the institutional level, communications with high authorities (HAs)

and list coordinators (LCs) will be primarily conducted through e-mail. Contingent upon the contact protocol to which the institution is assigned, HAs who are asked to send pre-notification e-mails, will be provided a tool that will make the process more efficient. At the individual level, communications will also be primarily through e-mail. Those who do not respond to initial e-mail contacts will be followed-up with through telephone calls and mail.

The Pilot ECD Survey respondents will have two methods by which to participate in the survey. The primary method will be via Web questionnaire. The secondary method is completing the Web questionnaire through a telephone interview, in which the telephone interviewer would access the respondent's survey, read the questions, and record his/her responses. Limiting survey participation to a single Web instrument reduces the costs of creating multiple versions of the questionnaire, improves data consistency and quality, eliminates costs associated with data entry, and reduces post completion processing time.

The data collection system used for Pilot ECD Survey will be designed so that respondents are able to change modes of completion should the need arise (e.g., a respondent begins the survey in the Web version, breaks-off, and upon reminder call asks if it is possible to complete the survey with the telephone interviewer picking up where they broke-off). The survey will have a real-time monitoring system, allowing NSF to monitor response status by mode, system problems, and comments from respondents. The survey will also provide the respondent with information and their progress through the instrument and send an automated e-mail confirming their completion and thanking them for their participation when they finish the questionnaire.

A.4 Identification of Duplication

Some overlap in terms of target population and content exists with the Survey of Doctorate Recipients (SDR) and National Survey of College Graduates (NSCG), however the Pilot ECD Survey will provide more comprehensive coverage of its target population (individuals working in U.S. academic institutions, FFRDCs, and NIH IRPs who earned their first doctorate within the last 10 years) than either the SDR or NSCG. The Pilot ECD Survey will also provide more complete data on foreign degreed doctorate holders.

While both the SDR and NSCG gather basic data about current employment and prior work experience, the Pilot ECD Survey collects substantially more information about respondent work experiences, especially with regard to postdoc experiences, professional activities and achievements, funding, and future career plans. As a result, the Pilot ECD Survey is the only comprehensive survey of recent U.S. and non-U.S. doctorate recipients working in the United States.

Within a sample drawn from the Survey of Earned Doctorates, the SDR is a survey of individuals through age 75 that earned their doctorate in a research field from a U.S. academic institution. Due to its frame, the SDR does not include doctorate holders currently employed in the United States who earned professional doctorates or earned doctorates from institutions outside the United States. Sample overlap with the NSCG is also expected to be minimal as only a small fraction of their sample will be ECDs.

The Pilot ECD Survey will attempt to contact all sample members, even those identified in the SDR sample, because the Pilot ECD Survey collects substantially more information. Of the 2,115 individuals sampled in the methodological study, only 41 (<2%) were in the 2013 SDR sample. The sample members identified in the SDR will receive a special letter describing their importance to the Pilot ECD Survey as well as the SDR.

A.5 Small Businesses Involvement

Not applicable. Small businesses will not be included in this round of data collection.

A.6 Consequences of Less Frequent Surveying

NCSES is seeking clearance for a single data collection cycle only for this pilot survey. In the future, NCSES envisions this survey transitioning to an every other year survey, alternating with and complementing the SDR data collection. NCSES also envisions a longitudinal component with follow-up interviews occurring approximately 5 and 8 years from doctorate award, and with the longitudinal respondents entering the SDR sample when they age out of the Pilot ECD Survey target population.

A.7 Special Circumstances

Not applicable. There are no special circumstances.

A.8 Federal Register Notice and Consultations outside the Agency

Federal Register Notice

The Federal Register notice was published on April 9, 2013 and amended on April 12, 2013 (see Attachment F). NSF received no public comment in response to the announcement by the closing date of June 10, 2013.

Consultations outside the Agency

NSF regularly consults with the Department of Education's National Center for Education Statistics (NCES) and other federal agencies, such as the NIH and the Department of Education, professional societies, such as the National Postdoctoral Association, and university staff. NSF staff members maintain frequent contact with members of the data-using community as well as with major academic data providers through attendance at professional society meetings and consultation with institutional and agency officials.

A.9 Payment or Gifts to Respondents

Not applicable. There are no payments to the Pilot ECD Survey respondents.

A.10 Assurance of Confidentiality

NSF and its contractors are fully committed to protecting the confidentiality of all survey respondents. The Pilot ECD Survey data will be collected under the authority of the National Science Act of 1950, as amended, and the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002. Institutional contacts (HAs and LCs) and prospective respondents are assured in contact materials that personally identifiable information (PII) and all responses will be kept confidential. Prospective respondents will also provide informed consent before beginning the survey. Attachment D (Confidentiality and Security Measures) outlines the other multiple steps during each stage of data collection that will be taken in order to protect both PII and survey responses.

Cover letters and survey questionnaire to each selected respondent advise them that the information they provide is confidential (see Attachment B – Institutional Contact Materials and Attachment C – ECD Contact Materials). The same notice of confidentiality (Attachment D) will be used in the introduction to the CATI interview as will be displayed prior to the start of the survey in the Web instrument.

Standard data collection procedures incorporate numerous safeguards for the data and must conform to a detailed security plan approved by NSF. All data collected and processed will be stored in secure internal data network with access available only to authorized users. While collecting the Pilot ECD Survey data, the PII data are separated from the survey response data and are not included in the analytic data sent to NSF. All project staff at the survey contractor and NSF will receive annual CIPSEA training to reinforce their legal obligations to protect their privacy and confidentiality of the Pilot ECD Survey and sign data use agreements annually to acknowledge this legal obligation.

The contractor takes special steps to ensure that data collected via the Web questionnaire are secure. First, access to the Web instrument is only allowed with a valid Personal Identification Number (PIN) and password correctly entered in combination. Second, data are transmitted by the Secure Sockets Layer (SSL) protocol that uses powerful encryption during transmission through the Internet. If a respondent keeps a Web survey open without any activity, the Web server closes it after a short period of inactivity, thus preserving the data up to the break-off point and securely closing the connection. The database architecture underlying the Web and CATI survey systems is designed such that authentication information and response data are maintained in separate tables, randomly generated identification numbers assigned for data collection purposes, and new, random identification numbers generated for the analytic files. This strategy provides an extra layer of security to protect respondents' data. Both development and production servers are backed up nightly, as required by the contractor's disaster recovery plan.

All published findings, analyses, and restricted databases are in formats which preclude individual identification and measures are taken so that the identity of individuals or organizations is not disclosed.

A.11 Sensitive Questions

The survey questionnaire does not contain any questions of a sensitive nature. While PII will be collected, those data will be omitted from all analytic files. Further steps to protect PII are described in Attachment D.

A.12 Estimated Response Burden

There are four types of respondents to the Pilot ECD Survey. At the first stage of sampling, the Pilot ECD Survey will select 201 institutions. At each institution, a high authority (HA) will authorize the institution's participation in the study and designate a list coordinator (LC) and an ECD contactor (EC). The LC will provide a list of all individuals working at their institution who earned their first doctorate or doctorate-equivalent degree within the past 10 years, including postdocs, nonfaculty researchers, and faculty members. Nonresponding institutions will be replaced with institutions of similar characteristics.

In the second stage, the EC will notify the individuals sampled from their institution of their selection and NSF will survey these sample members. To maximize response rates, we will ask the HA to serve as the EC or to designate another recognizable authority (e.g., Vice Chancellor for Research) to serve as the EC. The Pilot ECD Survey is intended to cover both U.S. and non-U.S. degreed and U.S. and non-U.S. citizens. The Pilot ECD Survey will sample 8,250 ECDs from 201 institutions. It is expected that 80% of the sampled ECDs will participate, yielding 6,600 ECD respondents.

The initial Pilot ECD Survey contact is made with the HA at each institution. The HA is asked to sign a study participation form, nominate a LC and an EC, and provide a letter of support for the survey. On average, HAs in the methodological study were able to complete these tasks in less than an hour. Assuming 201 HAs will participate, we estimate the total HA burden to be 201 hours.

The primary responsibility of the LC is to prepare a list of ECDs employed at the institution. In the methodological study, LCs (called POCs in the methodological study) required an average of approximately 6 hours to fulfill their duties. Assuming 201 LCs will participate, we estimate the total burden for LCs to be 1,206 hours. The ECs will send a pre-notification e-mail to sample members within their institution. In the methodological study, most ECs were able to complete this task in less than 30 minutes. It is expected that 5% of the sampled institutions will choose to have NSF send the pre-notification e-mails to ECDs. For this survey, we estimate a total burden of 96 hours for ECs.

Based on survey timing data from the methodological study and planned changes to the questionnaire, NCSSES estimates that ECDs will require 30 minutes on average to complete the questionnaire. Assuming 6,600 respondents, we estimate the total burden for ECDs to be 3,300 hours.

Taking into account all respondent types, we estimate the total respondent burden to be 4,803 hours. Table 1 shows detailed estimates of burden across respondent types.

Table 1. Pilot ECD Survey Estimated Burden by Stage and Respondent Type

Respondent Type	Minutes Per Respondent	Number of Respondents	Estimated Total Hours
<i>Stage 1: Frame Creation</i>			
High Authority (HA)	60	201	201
List Coordinator (LC)	360	201	1,206
Subtotal			1,407
<i>Stage 2: Individual Survey</i>			
ECD Contactor (EC)	30	201 (191)	96*
Early Career Doctorate (ECD)	30	8,250 (6,600)	3,300**
Subtotal			3,396
Total			4,803

*Assumes 95% of the institutions will send the pre-notification e-mails

**Assumes an 80% response rate

A.13 Cost to Respondents

Not applicable. This survey does not require respondents to purchase equipment, software, or services beyond those normally used in universities as part of customary and usual business.

A.14 Cost to the Federal Government

The total estimated cost to the Government for the Pilot ECD Survey is \$1.8 million for survey cycle costs and for NSF staff costs to provide oversight and coordination with other surveys. The cost estimated for the survey cycle is \$1.3 million, which is based on sample size; length of questionnaire; CATI and Web data collection technology; administrative, overhead, design, printing mail, and telephone data collection costs; data keying, editing, and processing; data quality control; imputation for missing item responses; weighting and estimating sampling error; file preparation and delivery; preparation of documentation and final reports; and survey evaluation. NSF staff costs are estimated at \$450,000 (\$150,000 annual salary of 1.5 FTE for 2 years of the Pilot ECD Survey cycle).

A.15 Changes in Burden

NSF expects that per respondent burden for the data collection will be similar to that observed in the methodological study. Total burden for the Pilot ECD Survey is expected to be 4,803 hours (see table 1 above). The substantial increase from the 1,754 burden hours in the methodological study is the result of increased sample sizes and higher expected response rates. In the methodological study, 1,519 or 71.8% of the 2,115 sample members responded to the survey invitation in some form. Given a longer data collection period and more intensive non-response follow up, we expect 80% of the 8,250 sample members to respond.

A.16 Project Schedule for Information Collection and Publication

The Pilot ECD Survey will be a two-stage data collection. In the first stage, sampled institutions are contacted to begin the frame building process. Because ECD listings from sampled

institutions will be received at various times, stage 2 data collection will be a rolling process beginning approximately one month after the receipt of the ECD listing from the institution and ending approximately two months later. As this is a pilot survey, NSF does not plan to use any complex analytical techniques in NSF publications using these data. The *InfoBrief*, detailed statistical tables, and public use data files will be published and made available on NCSES website. The time schedule for 2014 Pilot ECDS Survey data collection and publication is currently estimated as follows:

Table 2. Pilot ECD Survey Data Collection and Publication Schedule

Activities	Dates
Stage 1 (Institutional) Data Collection	June 2014 – October 2014
Stage 2 (Individual ECD) Data Collection	August 2014 – January 2015
Final Edited/Weighted/Imputed data file	April 2015
Public Use Data File and Documentation	June 2015 – September 2015
Survey Evaluation Report	September 2015

A.17 Displaying the OMB Expiration Date

The OMB expiration date will appear on all the Pilot ECD Survey webpages, including the welcome page, login page, confidentiality page, and survey questions pages. In the telephone follow-up it will be read to sample members during the introductory informed consent.

A.18 Exceptions in Item 19 on Form 83-1

Not applicable. There are no exceptions.