

NATIONAL SCIENCE FOUNDATION
SUPPORTING STATEMENT FOR PAPERWORK REDUCTION ACT
SUBMISSION
NATIONAL SCIENCE FOUNDATION
RESEARCH EXPERIENCES FOR UNDERGRADUATES PROGRAM

A. Justification

1. Necessity for the Data Collection

This study is being conducted by Mathematica Policy Research on behalf of the National Science Foundation (NSF) and in support of the Research Experiences for Undergraduate (REU) program. The REU program was created in 1987 to strengthen the science, technology, engineering, and mathematics (STEM) workforce. Building on research experiences as “one of the most effective avenues for attracting students to and retaining them in science and engineering, and for preparing them for careers in these fields,” the program is designed to foster student research and promote diversity (NSF 13-542). The program provides awards to principal investigators (PIs) at institutions—mostly universities and research centers—throughout the nation. Called REU Sites, these institutions engage groups of about 10 to 12 students in ongoing research programs or in research projects designed for the REU program, and operate mostly in the summer.

The main goal of the current study is to pilot test alternative approaches to collecting data *required by Congress in the America COMPETES Reauthorization Act of 2010*, which states that students in the REU program must “be tracked, for employment and continued matriculation in STEM fields, through receipt of the undergraduate degree and for at least three years thereafter” (Section 514[a][6] of Public Law 111-358). The legislation also mentions specific demographic characteristics of participants that need to be collected, such as gender, ethnicity, and enrollment in a two-year college. In addition to needing these data to report to Congress, NSF program officers (POs) and leadership need a more robust data system to enhance their efforts to monitor participation in the program and eventually to assess its effectiveness.

Data will be collected mostly through a newly designed REU Data System. The design of this system is based on (1) a careful review of the legislation (to ensure testing of a collection system that responds to the Congressional requirement) and the feasibility study conducted by the Science and Technology Policy Institute (Zuckerman et al. 2016); (2) existing REU Site proposals, reports, and current applications (to design a system that is tailored to the REU program); (3) the scholarly literature (to inform the general approach to data collection and enhance the NSF’s ability to use it in the future for evaluations); and (4) the Graduate Research Fellowship Program common application (GRFP; for benchmarking).

The present study includes designing the system and pilot testing it with a sample of REU Sites that volunteer to participate. By participating in this study, these Sites will have the opportunity to experience the data collections first hand and provide feedback that will be used to determine which approach will be the most effective, most efficient, and least burdensome for possible future implementation across all REU Sites.

Description of the Pilot

The pilot includes:

- a. **Testing two web-based approaches** to obtain basic student background and participation information (see the Appendix A for specifications):
 - **Registration.** The registration will be designed to collect the basic demographic and contact information needed for analysis and tracking purposes. Students wishing to participate in the REU program will be asked to register at a website through which they will submit information and obtain a unique ID. With this unique ID, they will then apply directly to the REU Sites using the existing Site application processes. Staff at REU Sites will later use the IDs provided by students to record application decisions and participation status of admitted applicants. Site staff will also be able to use the system to obtain data submitted by students who applied to their Sites.
 - **Common Application (including the Registration).** The common application will be modeled after existing REU Site applications and will replace those applications for the upcoming 2018 application cycle among participating Sites (for applications to participate in the program in 2019). The common application will enable students to apply to multiple Sites through one application.

Students will first complete the REU Registration described above, and then proceed to the common application through which they will submit additional information commonly required by Sites as part of their applications, such as transcripts. The system will also allow some level of customization for Sites to include additional requirements. Staff at REU Sites will use the system to provide information needed by potential applicants, retrieve applicant information, record application decisions and participation status among admitted applicants, and produce reports and run queries of data submitted by applicants to their Sites.

- b. **Obtaining and integrating educational and employment information.** The study will follow the subset of rising seniors who participate in the REU program in 2019 in the disciplines included in the pilot (rising seniors are the large majority of participants) to:
 - Obtain educational outcomes information from the National Student Clearinghouse (NSC)
 - Administer a survey to obtain information on employment outcomes (among those not pursuing studies at the time of the survey)

- c. **Conducting visits at a few REU Sites participating in the pilot** to interview principal investigators/coordinators and mentors and conduct focus groups with REU students. Conducted when most programs are operating in the summer (in this case, summer 2019), the site visits will be used to obtain in-depth feedback on the registration, common application, and data analytics tools available through the REU Data System for PIs.

The draft employment survey and site visit protocols (student focus groups, and PI/coordinator and mentor interviews) are included in Appendix B and C, respectively.

Description of Respondents

- **College/university students interested in applying to the REU program at the pilot Sites.** Students at two- and four-year colleges and universities are the target population of the REU program. Sites recruit incoming freshmen and rising sophomores, juniors and seniors. Through the registration or common application, these respondents will be asked to submit basic demographic and educational background information.
- **Reference letter writers.** Applicants in the common application will be required to provide contact information of two reference letter writers. The REU Data System will send reference writers an email requesting a reference letter in support of their application to the program.
- **Principal investigators (PIs) or staff designated by the PIs of participating Sites.** PIs or their designees will submit (1) information useful to prospective applicants (such as the opening and closing dates of applications) and (2) decisions/participation information (they will indicate which students were admitted and which later participated in the program at their Site). This information will be submitted through a module designed for PIs.
- **Former REU Site participants.** After students participate in the REU program, we will follow them to see if they can be found in the NSC and, if so, obtain educational participation and outcomes such as graduation from an undergraduate program and enrollment in graduate school. We will then seek to gather data on employment outcomes among those not enrolled in education programs or not found in the NSC.

2. Use of Information

The information collected through the pilot will enable NSF and other program stakeholders to assess these alternative approaches to collecting data needed to respond to the congressional requirement and meet the information needs of REU program officers and NSF leadership. These are:

- Registration + educational outcomes from the NSC + employment outcomes through a survey
- Common application + educational outcomes from the NSC + employment outcomes through a survey

These approaches will be analyzed based on several factors—including data quality, user burden, and user feedback—and compared against the current approach used by NSF to collect data.

3. Improved Information Technology to Reduce Burden

The pilot will create one online application to test both the registration and the common application. This online application will:

- Reduce application burden for students applying to the REU Program, as the same application can be submitted to multiple Sites (at present, most Sites require that applicants follow each Site's application procedures)
- Reduce burden on Site PIs and administrators who will no longer need to administer their own applications
- Provide analyses of applicant and participant data that is not readily available, is helpful to PIs, and is needed by NSF to monitor the program and comply with the America COMPETES Reauthorization Act; obtaining these data would otherwise generate additional burden for PIs and NSF staff

In addition, the pilot will include obtaining educational outcomes in a way that creates no burden for respondents (by relying on existing information available through the NSC).

4. Efforts to Identify Duplication

To minimize duplication, the registration and common application are designed to complement each other, including mutually-exclusive data items that are necessary for the pilot. In addition, PIs participating in the pilot will be asked to use the system to obtain the data they need, rather than requesting that applicants also submit the same information directly to the Sites. To this end:

- Registration. REU PIs participating in the registration pilot will get to preview the system before it is launched. They will also be informed ahead of time of the information to be collected through the registration to facilitate revisions to their applications and minimize the chances of collecting duplicate data.
- Common application. To ensure that the system captures the information they need and avoid concurrent application efforts led by program Sites participating in the pilot, PIs will be encouraged to preview the online application and customize it (1) to include additional questions they may wish to ask of applicants or (2) to request additional documentation not included in the common application already.

5. Efforts to minimize burden on small business

Small businesses are not affected by this information collection.

6. Consequences of Less Frequent Data Collection

This collection is needed to test alternative ways of collecting data needed to meet the congressional reporting requirements and monitor the REU Program. NSF commissioned a feasibility study to assess whether its current data collection approach would be adequate to meet the congressional requirements and consider alternative ways of collecting needed information. This feasibility study was conducted by the Science and Technology Policy Institute (STPI) and concluded that

“new data collection will be required, as the status quo of [REU] participants providing demographic information to NSF’s Research Performance Report System, coupled with voluntary tracking of participants’ career choices by the REU [principal investigators], is clearly insufficient to meet the [congressional] mandate” (Zuckerman et al 2016)

Based on an analysis of more recent cohorts of REU participants (other than those included in the STPI study) who responded to NSF’s new data collection through Research.gov (STPI analyzed data from earlier data collections through FastLane), Mathematica arrived at the same conclusion as STPI.

7. Special Circumstances

Not applicable.

8. Federal register announcement and consultation outside the agency

a. Federal Register announcement

A 60-day notice to solicit public comments was published in the Federal Register on February 21, 2018, at 83 FR 7498. No comments were received.

b. Consultation outside the agency

NSF commissioned a feasibility study to assess its ability to respond to the congressional mandate with its existing data collection efforts. As mentioned earlier, this study was conducted by STPI. NSF officers and Mathematica staff involved in this study reviewed the STPI report and attended their presentation at NSF—during which they were able to ask questions regarding issues raised and suggestions made in the report. An expert advisory board will also be convened to obtain feedback prior to launching the pilot.

9. Payment or gifts to respondents

Ten college students and five REU PIs will test the application before it is launched. The students will be offered a gift of \$40 for their participation. This amount is roughly equivalent to four hours of time at minimum wage—including about 2 hours to test the system, 1 hour debrief to share experiences using the system, and 1 hour for logistics

(coordination and review of background information).¹ PIs will not receive a gift for testing the system as they will be current beneficiaries of the REU program.

Other than payments to the ten college student application testers, no incentives or payments will be given to respondents.

10. Assurance of confidentiality

The collection of this information is a part of all regular applications to Sites participating in the REU Program. Applicants' information will be maintained in accordance with the requirements of the Privacy Act of 1974. No personal information will be released to the public. The system includes two notices:

The first one is based on the notice displayed in the NSF GRFP collection (OMB control number 3145-0023) and appears when users first access the system to obtain a user ID and a password. The notice reads as follows:

Notice

An agency may not conduct or sponsor, and a person is not required to respond to, an information collection unless it displays a valid Office of Management and Budget (OMB) control number. The OMB control number for this collection is XXXX-XXXX. Public reporting burden for this collection of information is estimated to average XX hours for registering and YY hours for registering and submitting an application, including the time for reviewing instructions. Send comments regarding the burden estimate and any other aspect of this collection of information, including suggestions for reducing this burden, to:

Suzanne Plimpton
Reports Clearance Officer
Division of Administrative Services
National Science Foundation
Arlington, VA 22230

Please note that information provided through the REU Data System will be used for admissions decisions, audits, and research and evaluation purposes. All applicants' information will be maintained in accordance with the requirements of the Privacy Act of 1974. No personal information will be released to the public.

The second notice appears when users certify and submit their registration or application. This notice reads as follows:

By clicking on the **SUBMIT** button below, I am certifying that the information provided is true and complete to the best of my knowledge. I understand that I am consenting to the

¹ Source of minimum wage information: <https://www.dol.gov/whd/minwage/america.htm>

confidential use of the information I provided for admissions decisions, audits, and research and evaluation purposes.

11. Justification for sensitive questions

Information regarding applicants' characteristics (such as gender, ethnicity and educational achievement) will be collected, as this information is needed by REU Sites to make admissions decisions and by NSF to report to Congress and monitor the program. Note that REU Sites normally collect this information as part of their own application processes.

Date of birth will also be collected to ensure proper identification of applicants and request educational information (enrollment and graduation information) from the NSC, which is important to comply with the legislative mandate while not increasing burden on participants.

Note that these data are also collected as part of the NSF GRFP common application (OMB control number 3145-0023).

12. Estimate of respondent burden

Table A1. Estimates of respondent burden by respondent type

| Category of Respondent | No. of Respondents | Participation Time | Annual Burden (hours) |
|--|--|---------------------------|------------------------------|
| Common Application and Registration Pilots | | | |
| 1. REU Applicants – Registration | 8,436 | .5 hours | 4,218 |
| 2. REU Applicants – Common Application | 6,865 | 12 hours | 82,380 |
| 3. Reference Writers– Common Application | 13,730 | .5 hours | 6,865 |
| 4. REU PIs (or their designated staff) who will complete and use the PI module | 263 | 8.9 hours | 2,341 |
| Site Visits** | | | |
| 5. PI/coordinator interview | 6 | 1 hour | 6 |
| 6. Mentors small group interview | 12 | .5 hour | 6 |
| 7. Student focus group | 36 | 1 hour | 36 |
| Educational Outcomes | | | |
| 8. This information will be obtained from the NSC, creating no burden for former REU participants | 1,578 (excluded from the total count) | 0 | 0 |
| Employment Survey** | | | |
| 9. REU Participant Employment Survey (estimate includes college students included in the field test) | 1,113 | under 15 minutes | 278 |
| Totals | 30,461 | | 96,130 hours |

Supporting information for Table A1

1. REU Applicants participating in the Registration

Number of respondents: Based on responses from NSF POs and REU PIs, we estimate that 145 REU Sites will participate in the registration pilot—121 Sites in Biology and 24 Sites in Earth Sciences. Based on past experience and existing NSF data, the NSF REU Program Coordinator estimates that, on average, students apply to 3.3 Sites and Sites in these disciplines receive 192 applications. We therefore estimate 8,436 unique system users in the registration pilot (192 applications per Site multiplied by 145 Sites and divided by 3.3 to identify unique applicants).

Participation time (number of burden hours per respondent): We estimate that the subset of 26 mostly multi-choice questions included in the registration would take a respondent about .5 hours to complete, assuming about 1 minute per question and 4 to 5 minutes to answer a handful questions needed to obtain a username and create a password.

2. REU Applicants participating in the Common Application

Number of respondents: Based on responses from NSF POs and REU PIs, we estimate that 118 REU Sites will participate in the common application pilot—98 Sites in Engineering and 20 Sites in Mathematics. Students applying to these Sites will complete the questions included in the registration before proceeding to additional items needed to submit their full applications through the common application. Based on past experience and existing data, the NSF REU program coordinator estimates that, on average, students apply to 3.3 Sites and Sites in these disciplines receive 192 applications. We therefore estimate 6,865 unique system users in the common application pilot (192 applications per Site multiplied by 118 Sites divided by 3.3 to identify unique applicants).

Participation time (number of burden hours per respondent): Based on the NSF GRFP common application, we estimate that it will take 12 hours to complete the full application.

3. Reference Writers– Common Application

Number of respondents: Students using the common application are required to submit contact information for two reference writers. The REU Data System contacts reference letter writers by email. Reference writers will be asked to complete only one form for each applicant that names them as a reference, regardless of the number of Sites to which students apply. We estimate 6,865 applicants in the common application and, therefore, at most 13,730 reference writers.

Participation time (number of burden hours per respondent): We estimate that it will take 0.5 hours for reference writers to draft a short paragraph about the applicant and rate the applicant on 11 characteristics (by selecting check-boxes).

4. REU Principal Investigators (PIs) in Registration and Common Application

Number of respondents: We estimate that a total of 263 PIs or their designees—one for each Site—will provide information through the REU Data System that will house the registration and common application. This includes 121 Sites in Biology, 24 in Earth Sciences, 98 in Engineering, and 20 in Mathematics.

Participation time (number of burden hours per respondent): We also estimate that the PI burden is 8.9 hours for each Site. This number comes from the time it takes PIs to:

- a) Provide information about their Site and submit additional application requirements (24 items at a rate of 1.25 minutes per item = 0.5 hours per Site)
- b) Record applicant IDs and admission and acceptance decisions for Sites in the registration (3 items per each of the 192 applicants at a rate of 1 minute per item = 9.6 hours per Site)
- c) Record admission and acceptance decisions for Sites in the common application (2 items per each of the 192 applicants at a rate of 1 minute per item = 6.4 hours per Site)
- d) Record participation information (1 item for 12 participants at the rate of 1 minute per item = 0.2 hours)

Weighing the estimates in b) and c) by the number of Sites piloting each approach to data collection (145 Sites for the registration and 118 for the common application)

5-7. Site Visits

Number of respondents: We estimate a total of 6 PIs and 12 mentors will be interviewed (two per Site) and 36 students will participate in focus groups (6 per Site).

Participation time (burden hours per respondent): We will conduct one-hour interviews with PIs/coordinators and one-hour focus groups with students. Interviews with mentors will last 30 minutes (if individual) to an hour (if small group).

8. Educational Outcomes

Number of respondents: We will obtain educational outcomes data from the NSC. We will request data for individuals who (1) participated in the common application or registration pilots, (2) were selected and participated in the REU program, and (3) were rising seniors at the time of their participation. We assume that approximately 10 students participate in each of the 263 Sites in the pilots and nearly 60% of them are rising seniors (based on STPI's analysis), resulting in a total of 1,578 students ($10 \times 263 \times .6$).

Participation time (burden hours per respondent): Not applicable, as the data are available through the NSC.

9. Employment Survey

Number of respondents: The employment survey will be administered to students who (1) participated in the common application or registration pilots, (2) were selected and

participated in the REU program, (3) were rising seniors at the time of their participation, and (4) are not enrolled in further education the year after obtaining a bachelor’s degree. Given that the program seeks to serve 10 to 12 per Site, and some Sites enroll 8 to 10 students, we assume that approximately 10 students participate in each of the 263 Sites in the pilot, nearly 60% of them are rising seniors (based on STPI’s analysis), and about 70% are not enrolled in school.² In addition, 8 students will field test the survey (2 in each participating discipline). This results in a total of 1,113 survey respondents $((10*263*.6*.7) +8)$.

Participation time (burden hours per respondent): The survey will take no more than 15 minutes to complete. This time should be sufficient to answer the few questions needed to comply with the congressional requirement, plus a few questions needed to meet the information needs of NSF leadership and program officers. We will seek to keep the survey short to avoid unnecessary burden and improve response rates and data quality. Note that the literature does not indicate a strong correlation between survey length and response rates. However, research does suggest that “stated” survey length of time is associated with lower response rates, and survey length influences data quality of questions positioned later in the survey (Galesic and Bosnjak 2009).

13. Cost Burden to Respondents

There are no direct costs to respondents. The cost to respondents involves solely the time spent completing their REU applications (which they would incur in the absence of this pilot) and any follow-up surveys.

14. Cost Burden to the Federal Government

The estimated cost of this data collection is \$980,562, which includes the design, development, and implementation of the REU Data System for the 2019 REU cycle.

15. Reason for Change in Burden

Not applicable. This is a new information collection.

16. Schedule for information collection and publication

| Task | Date |
|---|--------------------|
| REU Registration and Common Application Pilot | August 2018 |
| Site Visits | Summer 2019 |
| Interim Report | Winter 2019-2020 |
| National Student Clearinghouse | Fall 2020 |
| Employment Follow-up Survey | Spring-Summer 2021 |
| Final report | Winter 2021 |

² Available estimates indicate that, among bachelor’s degree recipients, about 30% enroll in further education within a year of completing their graduate degree (U.S. Department of Education 2016) and 39% enroll in a graduate program within four years of graduating from college (Baum and Steele 2017).

17. Display of OMB expiration date

The expiration date for Office of Management and Budget (OMB) approval will be displayed as shown in section 10 of this document.

18. Exception to the certification statement

There are no exceptions to the certification statement.