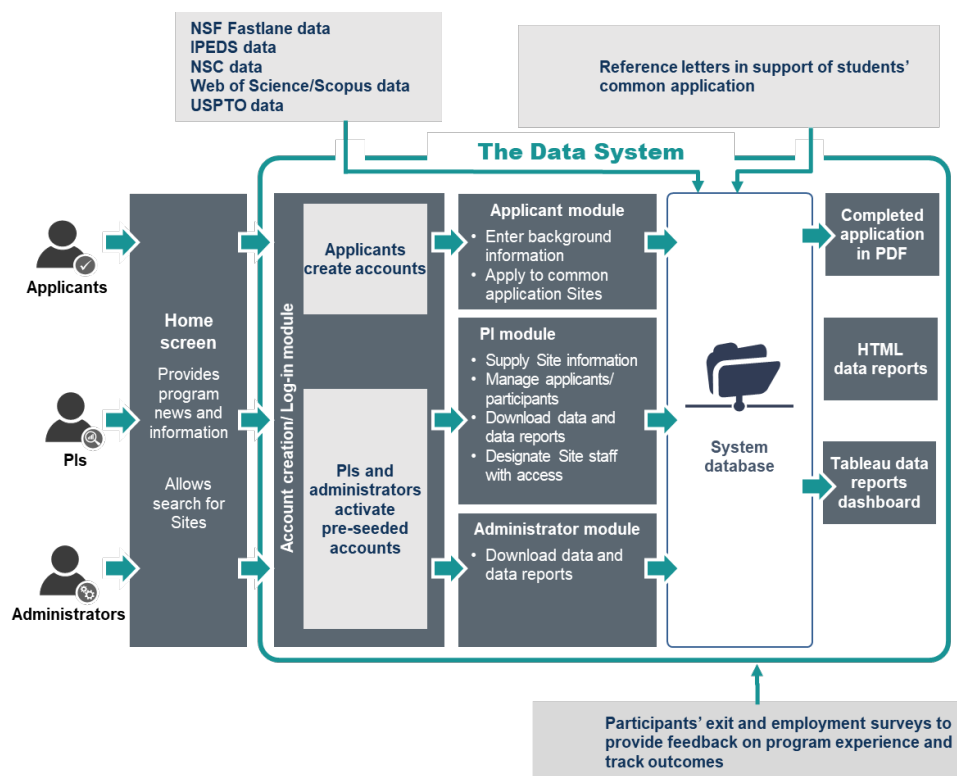# Attachment A ● Data collection overview

Data collection overview

The **Education and Training Application (ETAP) system** will be an online platform designed to offer NSF a robust and secure data collection system and data enclave for the IRES and REU programs, and other programs that similarly invest in training opportunities for students. Based on its predecessor (the REU data system), the system will serve as a centralized location where student applicants can provide information required to participate in NSF programs, PIs can access applicant information and submit information about the training opportunity they offer, and NSF staff can access information about applicants, participants, and opportunities.

1.1. System modules

The **data system** will be the restricted-access portion of a public website and will include four functional modules: (1) account creation/log-in, (2) applicant, (3) PI, and (4) administrator modules—along with a Tableau for interactive reporting of system data (Exhibit 1).

Exhibit 1. The ETAP system at a glance

IPEDS= Integrated Postsecondary Education Data System
NSC= National Student Clearinghouse data
USPTO= United States Patent and Trademark Office

An **account creation/log-in module**, through which users can receive an ID and establish an email to access the system and reset their password

An **applicant module**, through which students can enter information necessary to apply to a NSF opportunity through a common application or register to participate in a program. Both require applicants to complete a registration form designed to collect basic demographic and contact information but differ in what happens after the registration is completed.

> **Registration with common application.** Applicants applying to training opportunities using the common application will be able to apply to multiple opportunities by completing two application components. Applicants will first complete basic registration information and then proceed to the additional opportunity requirements, through which they will submit transcripts and any additional information commonly required by Sites.

> **Registration**. Applicants registering to participate in an NSF opportunity that is not running competitive applications will only need to register through the system without any additional requirements.

A **PI module**, through which PIs can enter information about their project (opportunity offered), access application materials, submit admissions decisions and participation information, and access data reports

An **administrator module**, through which NSF program officers can access aggregate data reports and system administrators can access applicant-level data.

The database will store information entered by the users who access the front-end modules and will integrate that information with data obtained from other external sources. Unless otherwise noted, beginning fall 2021, the database will include the following:

- Award information retrieved from the NSF FastLane system to reduce burden on PIs, such as the award title and institution address
- Opportunity information provided by PIs through the data system, such as expected program dates, location, and opening and closing date for receiving applications
- Student information provided by applicants through the registration (in the future) or common application
- Information on applicants' background and aptitude provided by reference letter writers
- Information on applicants' home institutions retrieved from the Integrated Postsecondary Education System (IPEDS) and the Carnegie Classification of Institutions of Higher Education to facilitate analyses, such as calculating the proportion of participants who come from a two-year college or an institution classified as having "very high research activity"
- Information on participants' research experiences (activities and mentorship received), learning gains, and satisfaction—provided by participants at the end of the research experience
- Information on participants' educational outcomes—to be obtained from the National Student Clearinghouse (NSC) after participants' expected graduation dates (to be included in the 2023)
- Information on participants' employment outcomes—to be obtained from those participants not enrolled in undergraduate or graduate programs according to the NSC data (to be included in the 2023)

- Information on participants' research productivity outcomes—to be obtained from administrative data on publications and patents through name-matching (to be included in the 2023)

1.2. Key system features

The data system is envisioned as a robust and reliable data collection system that accommodates the needs of different users. When designing the system's features, NSF prioritizes the goals of collecting and disseminating useful information, ensuring data quality, reducing burden, and providing a positive user experience. Next, we describe the key system features.

Features for applicants

Includes an opportunity search tool. The data system provides users with a list of NSF-funded opportunities that are using the system along with information that PIs provide (such as program dates and the application deadline). The list includes a search tool that enables users to filter opportunities by characteristics such as program type (for example, IRES, REU), student type (for example, undergraduate and graduate), Site/Project location and discipline.

Key system features

For applicants

– Includes an search tool for NSF-funded training opportunities
– Accepts only applications from eligible students
– Allows restrictions to the maximum number of applications (which could be submitted at different times)

For PIs

– Flexible and customizable
  • PIs choose whether to run applications competitively or non-competitively
  • PIs control when to open and close applications to their award
  • PIs can add questions and request additional materials*
  • PIs can authorize system users (e.g., co-PIs)
– Electronic reference letters*
– Two options to report admissions decisions and participation information (interactive or bulk upload)
– View, print, and download data
– Data reports

For NSF program officers

– Data reports (including a Tableau dashboard)

*Common application

Limits the number of applications students can submit (through the common application). The system will be programmed to restrict the number of applications a student can submit. The limit was up to 10 applications per student in the REU data system and can be modified based on program needs.

Has built-in features to enhance data quality, such as required fields, validation checks, and programmed skip logics. Nearly all fields in the system require responses, but for items with voluntary responses, such as those that ask about demographic information, applicants may indicate that they do not wish to provide the information requested. In addition, responses are subject to edit checks to ensure applicants enter all the required information, and the information is valid. For example, test scores must be submitted within a valid range. Email addresses and dates must follow a particular format. Validation checks also prevent applicants from submitting applications with missing responses or invalid values and prompt them to fix any issues. Lastly, conditionally required responses are integrated into the flow of questions, such as prompting for more information if an applicant selects "other" in response to certain questions.

Implements predictive search for college names and uses standardized college majors. The system incorporates the 2017 list of colleges from IPEDS, which activates by way of predictive search once users start typing the name of the institution of current enrollment. The system also collects information on fields of study consistent with the National Survey of College Graduates.

Presents progress status and review panel page. The applicant module includes a reporting page that provides at-a-glance status reports of completion progress by indicating the sections of the application that are still in progress. Before submitting the application, users can review it for accuracy in print mode.

Provides capabilities to upload and download PDFs. The system enables applicants to upload PDF files with their transcripts, curriculum vitae, and other documentation, all of which are accessible to PIs of the opportunities to which they have applied. In addition, students can view (and download) PDF documents that PIs may have loaded in the system to describe additional program requirements.

Features for PIs

Offers flexibility and customization for PIs to manage their applications. For example:

- **PIs decide how to run applications to their award.** PIs determine whether to run applications by prospective participants competitively and non-competitively.
- **PIs control the application period (common application).** PIs determine the dates on which they wish to start and stop receiving applications. PIs can modify the dates at any time.
- **PIs can add questions and request additional materials (common application).** PIs can upload a PDF and add up to three questions to the application. The added requirements are administered to students who attempt to apply to that PI's Site/Project.
- **PIs can authorize users to manage their Site/Project in the system.** PIs can designate co-PIs and staff members with similar access rights who can manage the Site/Project account.

Offers two options to report admissions decisions and participation information (interactive or bulk upload). PIs can report admissions decisions and acceptances interactively in the website for each applicant by using drop-down menus, or they can upload a CSV file with the information.

Supports electronic reference letters (common application). Applicants are required to provide contact information for two people who will provide a letter of reference. The system then submits

an email inviting those people to complete an online form about the applicant. PIs can access all applicant information, including submitted reference letters, electronically in one system.

Supports view, print, and download data functionalities. PIs can view and print an applicant's completed information in a review screen. PIs can also download applicant data for their own applicants, along with admissions and participation decisions, in a CSV format. Filtering tools are available to users who wish to limit the scope of downloaded data based on a variety of factors, such as applicant demographic characteristics or the type of institution attended by applicants. Note that the database is configured to allow PIs access to data on applicants and participants based on their data access privileges. (PIs can only access applicant-level data for those who have applied to their project.)

Pre-loads data from FastLane to enhance data quality and reduce burden. Award information available in FastLane (which includes the PI contact information and the award number, discipline, title, and abstract) are loaded into the system. Once PIs submit their grant award number in the PI module, the system populates information based on the grant award ID. This feature reduces the burden on PIs—who do not have to submit information that has already been provided to NSF— and enhances data quality, because PIs can review the information and request revisions as needed.

Presents progress status. The PI module also has a summary page that provides at-a-glance status reports of their application cycle, including a list of authorized users, number of applications received, and progress on making admissions decisions.

Provides reports on aggregate descriptive statistics. PIs and their authorized users have access to aggregate descriptive statistics (in the form of HTML, PDF, and CSV tables) for their own award, their own discipline, and the program as a whole.

Features for administrators

Provides reporting and data analytics. NSF program officers and other administration users will have access to a Tableau dashboard with which they can generate descriptive statistics for different samples of the data (for a given award, a given discipline, or the program as a whole).

Allows data download. This functionality has been built and, at present, is restricted to authorized system administrators. In consultation with NSF's office of general counsel and other key NSF staff, NSF will determine, and develop procedures to support, data access and use.

Features for all users

Sends automated email notifications. The system will send automatic emails through the website domain in support of communications, to include: (1) a welcome email that confirms users have successfully created an account; (2) emails allowing users to reset their password; (3) emails to people chosen to write letters of reference, with a secure link for them to upload their letters (common application only); (4) emails inviting users to create an account when the PI authorizes them to manage their account in the system; (5) an email confirming successful submission of the application form; and (6) an email administering the participant exit survey.

1.3. Data elements

The system will provide a common data collection infrastructure that serves NSF needs of consistency and standardization across programs, while allowing flexibility for PIs to request additional information needed for their operations. The data system will allow users to collect basic information needed to monitor the characteristics of participants (such as gender, ethnicity, education level, discipline, and country of experience) and learn about their experiences participating in the programs. Through this system, critical data will also be gathered to facilitate future efforts to obtain information on participants' educational outcomes through the National Student Clearinghouse (at low cost and no burden to participants), administer surveys to gather employment outcomes, and evaluate the program.

Applicant data. Exhibit 2 shows the data elements planned to be collected from applicants through the system (the data collection instrument is available in Attachment B).

Exhibit 2: Data elements that will be collected from applicants through ETAP
(*) Section required from all applicants.

Section	Data elements	
Create an account/log-in (*)	**Personal information** • Name • Preferred name • Preferred pronouns • Date of birth • Email address	**Data sharing preferences** • Preference for data sharing with other NSF opportunities for which they have not applied
Registration (*)	**Contact information** • Cell phone number • Alternative email address • Mailing and permanent address • Social media account information **Demographic information** • Race • Ethnicity • Sex at birth • Gender identity • Veteran status • Disability status **Family** • Education and occupation of applicant's parents or guardians **Education** • High school • Prior postsecondary degrees • College entrance test scores	**Current enrollment** • Name of college or university • Enrollment status • Degree pursued • Primary field of study • Expected completion date • Class standing • College GPA • College financing **Prior experience** • Prior research experience (NSF or not NSF-funded) • Prior international research experience **Work experience** (if applicant is a teacher) • School name • Teaching grade and subject • Years of teaching experience • Teaching certification **Program awareness** • Awareness of the program **Opportunities** Opportunity applying for
Additional Application Requirements	**Other application materials** • Personal statement • Transcript • CV or resume • References' contact information • Additional materials requested by PIs	

Section	Data elements
Certify & Submit (*)	**Certify and submit**

Participant data. In addition to baseline (application) data, the system collects detailed information from participants. For example, the participant exit survey (as designed for the REU and IRES programs) is administered through the system and includes questions to gauge participant satisfaction with the research experience, support received in preparation for and during the experience (such as scientific preparation and more practical training to foster a successful adaptation in the foreign research environment), mentoring received (and by whom), collaborations formed, skills gained, and so on. The exit survey includes 15 questions common for participants in either program with 2 additional questions for participants with an international program experience (Attachment B includes the survey instrument).

1.4. Data reports

The data reporting capabilities of the system differ by type of user, as shown in Exhibit 3.

Administrators. NSF program officers have access to **data tables** that show aggregate statistics of the data collected. These tables will be refreshed every morning at 4 a.m. EST to provide updated ("live") data. Once program officers log in, they can:

Click on the table name to view the table in HTML

Click on a PDF icon to print or save the table as a PDF

Click on a CSV icon to download the table as a CSV file

In addition, NSF will access to a **Tableau dashboard** that shows aggregate statistics of the data collected in an accessible and visually appealing way. Program officers using the dashboard can interactively choose the data displayed based on samples (for example, all awards in the program combined, or a particular award).

Lastly, NSF program officers can request access to the **applicant-level data** for monitoring, research, and evaluation. The data come in one CSV file with one record per applicant.

PI (and their designated users) will have access to data tables showing the same data elements as those in the administrator module but for different samples of the data. Data tables for PIs show aggregate reports on their own award, their own discipline (if applicable), and on all awards in the program combined. In addition, PIs will have unrestricted access to applicant data of applicants to their own opportunity. Due to software licensing restrictions, PIs will not have access to the Tableau dashboards.

Exhibit 3. Summary of the system's data reporting capabilities

	Data tables	Tableau dashboard	Applicant-level data
NSF program officers	✓ Aggregate reports on: • Each discipline (if applicable) • The program as a whole	✓ Aggregate reports on: • Each award • Each discipline (if applicable) • The program as a whole	✓ Disaggregated data for program as a whole in CSV file (upon request)
PIs and designated users	✓ Aggregate reports on: • Their own award • Their own discipline (if applicable) • The program as a whole		✓ Disaggregated data for their own Site in CSV file (available through the system)