



ECR Hub
EDU Core Research



Secondary Data for STEM Education Research

An Overview

Dr. Terri Pigott | Dr. Emma Cohen | Dr. Debshila Basu Mallick

April 2, 2026



Virtual Meeting/Conference Recording Notice

- The American Institutes for Research® (AIR®) allows for the recording of audio, visuals, participants, and other information sent, verbalized, or utilized during business-related meetings. By joining a meeting, you automatically consent to such recordings. Any participant who prefers to participate via audio only should disable their video camera so only their audio will be captured. Video and/or audio recordings of any AIR session shall not be transmitted to an external third party without the permission of AIR.
- In addition, AIR does not permit participation in AIR meetings through the use of AI bots such as Otter.ai or other AI platforms to record or transcribe conversations for AIR meetings, webinars, or virtual events in lieu of in-person attendance, unless requested as a reasonable accommodation. Any participants who attempt to use AI software for meeting participation will be denied admission, or their session will be terminated once it becomes apparent such software is in use.



Meet the Presenters



Dr. Terri Pigott

ECR Hub Co-PI
*Professor, Georgia State
University*



Dr. Emma Cohen

ECR Hub Project Director
*Senior Researcher,
American Institutes for
Research*



Dr. Debshila Basu Mallick

Scientific Director
Safelnsights



About ECR:Core and ECR:BCSER solicitations

The **EDU Core Research (ECR) program** at NSF supports **fundamental STEM education research and capacity-building projects** to conduct such research, complementing the more applied focus of other NSF EDU programs:

- Fundamental research projects through the **ECR:Core** solicitation
- Capacity-building projects through the ECR Building Capacity in STEM Education Research (BCSER) solicitation



Still Active Parts of ECR:Core

Present status:

- ECR:BCSER is currently archived and no longer accepting proposals.
- ECR:Core solicitation is still active (next deadline of Oct 1, 2026).
- Still accepts proposals for research focused on undergraduate education, graduate education, higher education and the STEM workforce.



What is Fundamental Research?

- Fundamental research generates knowledge and understanding with the potential for broad relevance (**theory building**).
- Implications for improving STEM education practice may be **indirect and long-term rather than direct and immediate**.
- See more here: <https://ecrhub.org/what-ecr-funds>



About ECR Hub

Two overarching goals of the ECR Hub

- **Extend the influence and reach** of existing ECR:Core and ECR:BCSER investments
- **Build career development opportunities**, especially for underrepresented institutions and scholars

Learn more about ECR projects

- **Projects library:** filter through ECR projects
- **Publications library:** publications related to ECR projects
- **Impact stories:** how ECR projects impact the field



Overview of webinar

- Many in the ECR grantee community are wondering how to continue their research agendas in a changing funding landscape
- This session is designed to share some ideas about how to continue, extend, or pivot your research agendas using already available data
- We offer a jumping off point
 - We will discuss data resources you may or may not be aware of already
 - This is not a comprehensive list of available datasets, nor a technical training on how to use each one
 - Our focus will be on quantitative data sources, since those are most readily available



Learning Objectives

1. **Understand** what secondary datasets are and how they can benefit your research
2. Review examples of **research questions** that can be answered with secondary data relevant to fundamental STEM education research
3. Explore different **sources** of secondary data
4. Learn about important **considerations for working with secondary data**
5. Review different ways to **access secondary data** and seek training and support

This community already has valuable knowledge – please share resources and ideas in the chat throughout the session!



What is secondary data? How can it benefit your research?



Secondary data are...



Data that were collected by **someone else**

For a purpose **that is not your current purpose**

But can be used to **answer additional questions**

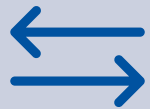


Why consider working with secondary data?



Readily available

You don't need to wait to collect data – you can just start cleaning and analyzing



Breadth and scope

Secondary data may offer more information than an individual investigator can typically collect on their own, e.g., nationally representative data or “big data”



Cost

While some datasets are free and others have licensing charges, using existing data is typically less expensive than conducting a new data collection

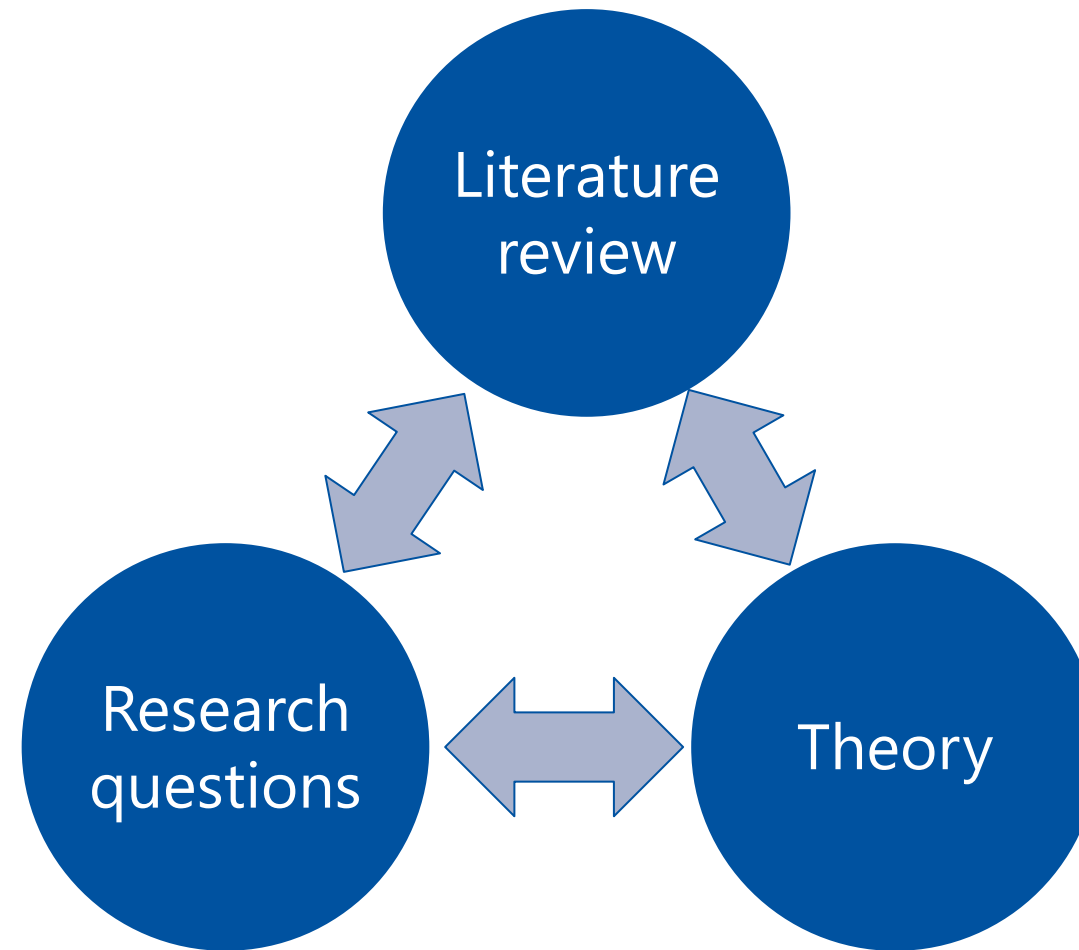


How do I get started?

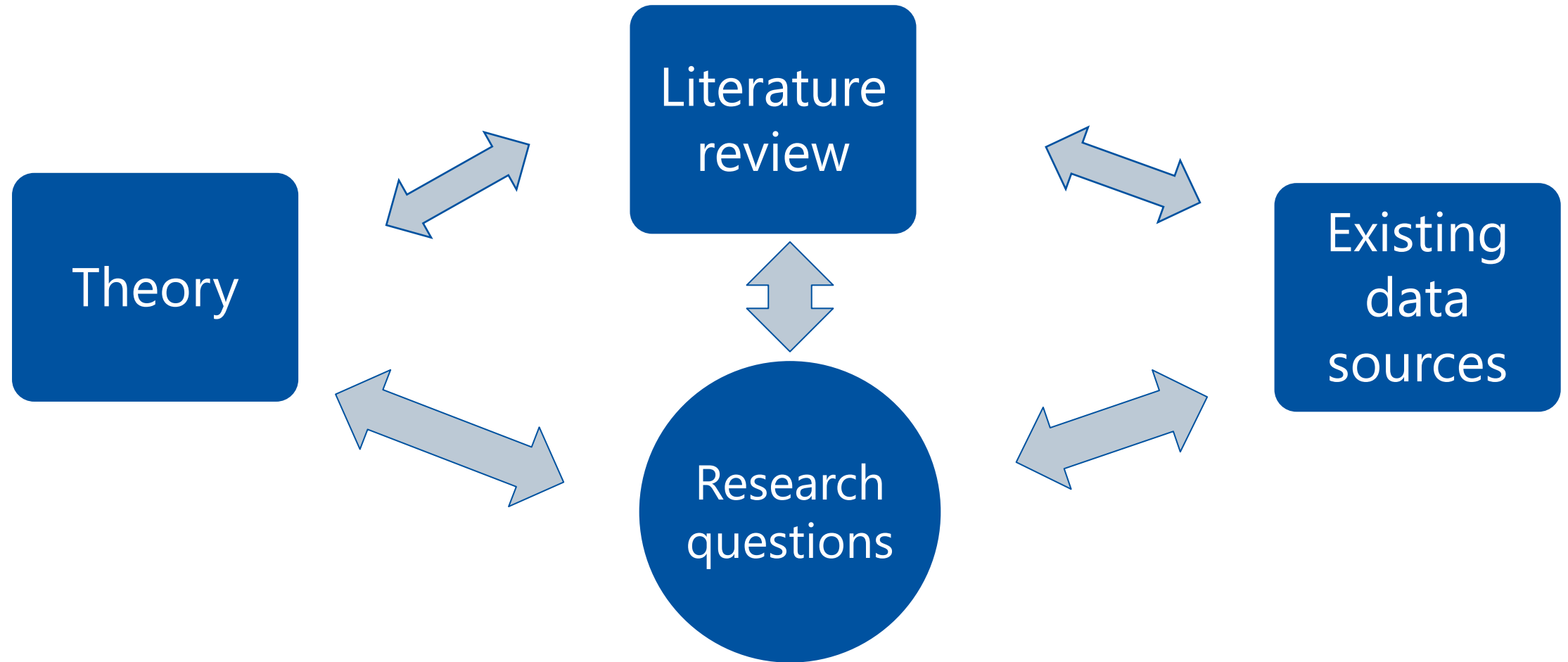
- Just like with primary data collections, secondary data-driven research starts with a research question.
 - BUT: the data sources themselves can be a consideration in the research questions.



Paradigm for developing RQs



Paradigm for developing RQs for a secondary study



What questions can fundamental STEM education researchers answer with secondary data?



Before we look at some specific topics...

- Be aware that some of the topics we discuss in this section are ***not fundable through ECR:Core based on the current solicitation.***
- We are including them anyway, because we expect they are of interest to our audience, and because the goal of this session is to spark paths forward for new research with less and/or different funding
- We are going to look at some specific questions, as a way to think about what *kinds* of data are available.



Example research questions and topics

Setting	Research on STEM Learning and Learning Environments	Broadening Participation in STEM	Research on STEM Workforce Development
K – 12 education	<i>We will hear more about this from SafeInsights later in the presentation</i>	Which factors are associated with change in math self-concept during elementary school, and how does this vary by race/ethnicity and socioeconomic status?	What types of coursetaking patterns in early high school are most associated with majoring in STEM in college?



Example research questions and topics, continued

Setting	Research on STEM Learning and Learning Environments	Broadening Participation in STEM	Research on STEM Workforce Development
Undergraduates	How does participation in high impact practices differ between STEM and non-STEM majors?	What factors are associated with underrepresented groups majoring in STEM?	What kinds of occupations do undergraduate STEM majors achieve, and how does this vary by discipline?
Graduate students	<i>Possible gap</i>	What factors are associated with women graduate students going onto a postdoc?	How do doctoral training experiences relate to sector of employment, job-field match, and early career trajectories?



Example research questions and topics, continued

Setting	Research on STEM Learning and Learning Environments	Broadening Participation in STEM	Research on STEM Workforce Development
Higher education	How do institutional characteristics relate to STEM completion rates?	Which institutional contexts are associated with smaller or larger equity gaps in STEM retention/completion?	How do universities' research and training investments align with regional STEM workforce needs?
STEM Workforce	<i>Possible gap</i>	Which STEM majors have more equitable funneling of students into STEM occupations?	How are skill requirements changing for STEM occupations?



Example research questions and topics, continued

Setting	Research on STEM Learning and Learning Environments	Broadening Participation in STEM	Research on STEM Workforce Development
Informal learning environments	How do informal learning opportunities such as home science and math activities or involvement in different enrichment activities relate to math achievement?	How does access to informal learning opportunities vary by race/ethnicity, gender, and socioeconomic status?	<i>Possible gap</i>

REVISE center hosts a [repository of AISL projects](#), some of which link to websites where de-identified data can be accessed.



Exploring different sources of secondary data



SafeInsights



Bringing ECR questions to SafeInsights

Debshila Basu Mallick
Scientific Director

April 2, 2026

Today's Agenda

1. A little bit about SafeInsights

Research model

Data ecosystem

2. ECR Core & SafeInsights' mutually aligned priorities

3. Progress to date

01.

A little bit about SafeInsights

Safelnsights is a major education research infrastructure

Accessible via Safelnsights Research Portal



Technical

3 analytical frameworks to support large-scale R&D, AI-augmented supports for coding, testing, and review



Logistical

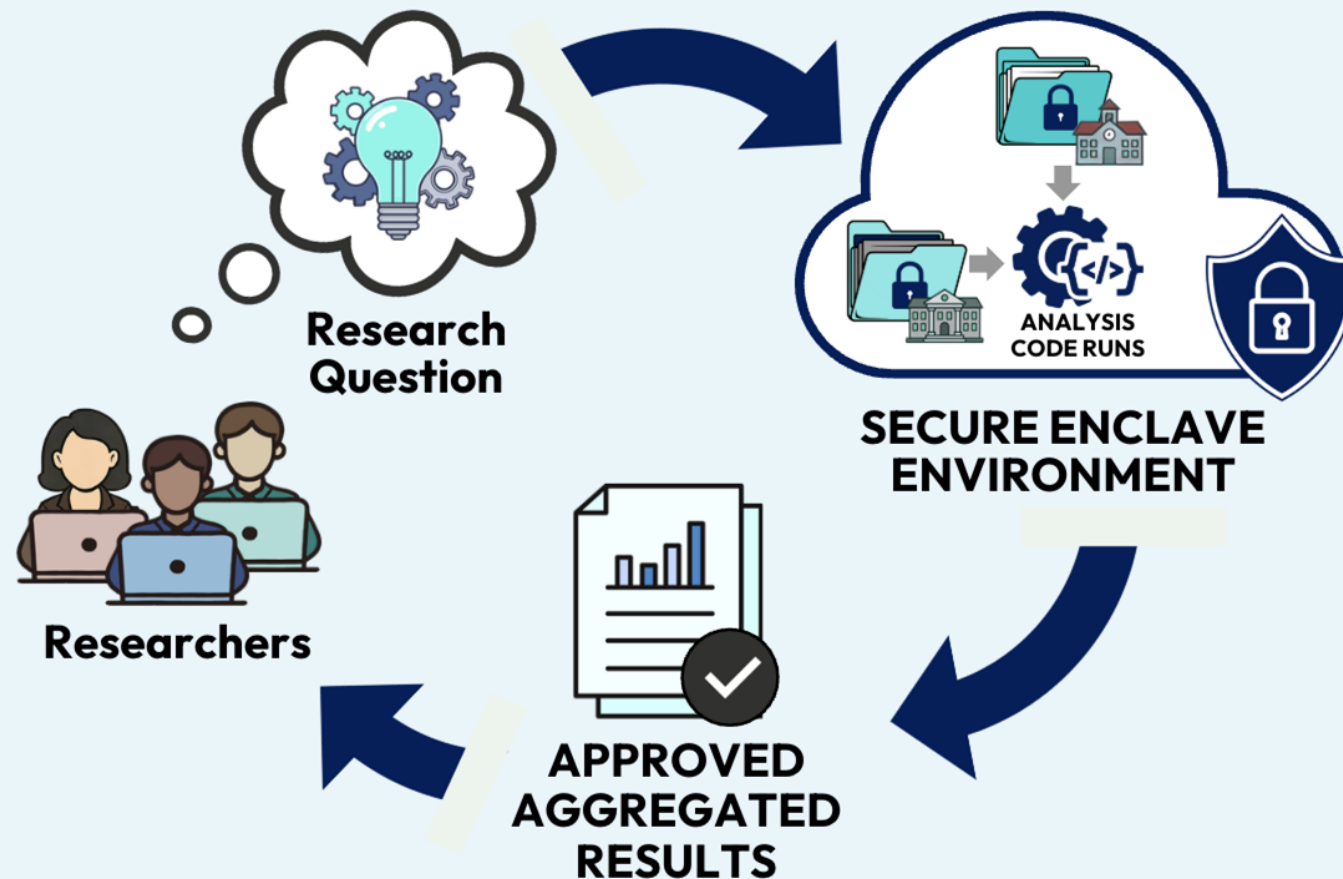
Data Sharing Agreements,
IRB, Governance,
Sustainability,
...



Community

Make research work for
schools, students,
researchers, edtech

SafeInsights flips the traditional research model



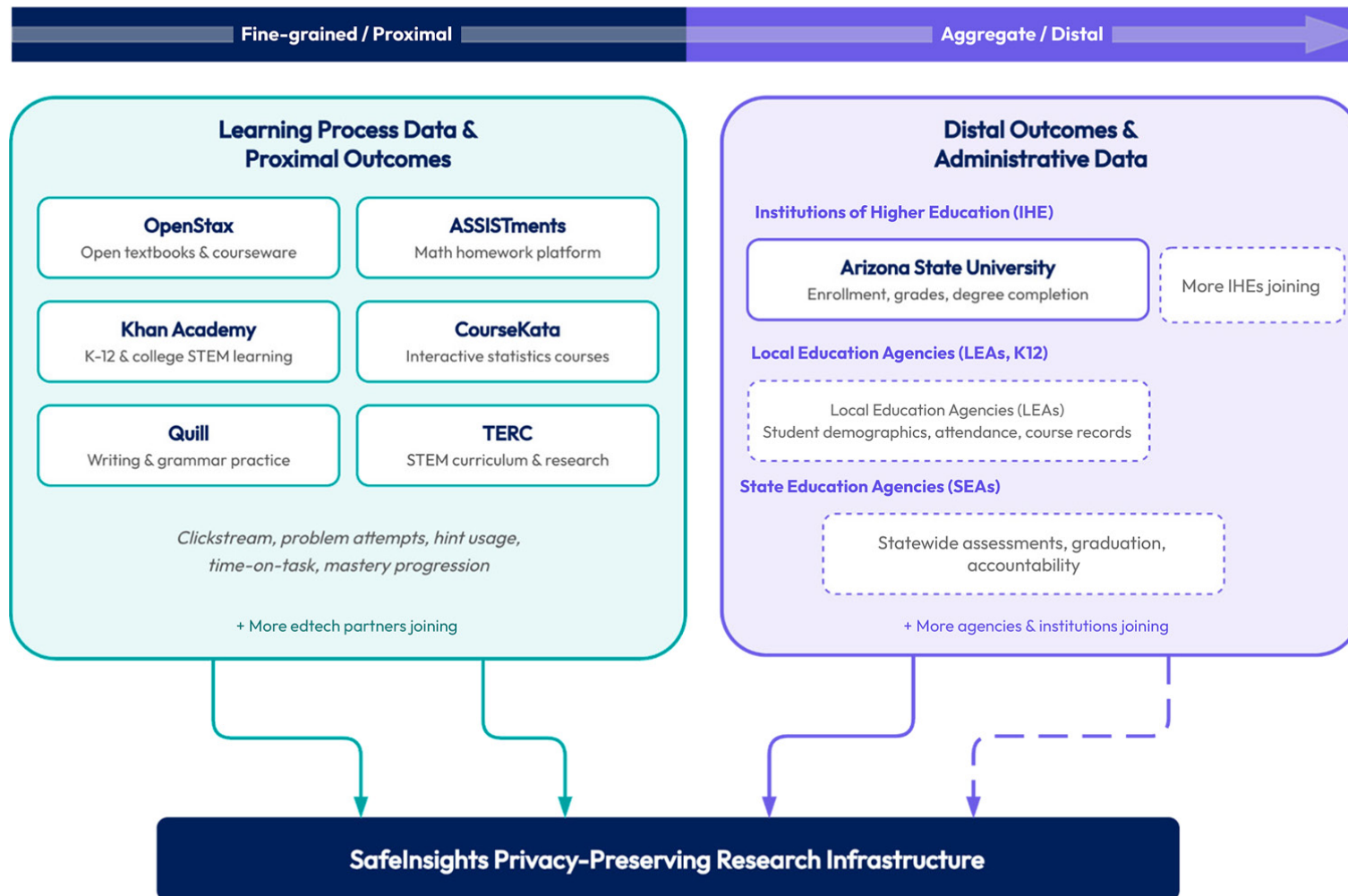
SafeInsights presents research questions to the data, preserving privacy

Our Mission

**To advance research
that drives better learning
for everyone.**

From fine-grained learning process to educational experience and outcomes

SafelInsights Data Partner Ecosystem



Growing Ecosystem: Connecting Diverse Data Sources for Fundamental STEM Education Research

02.

ECR Core and SafeInsights' mutually aligned priorities

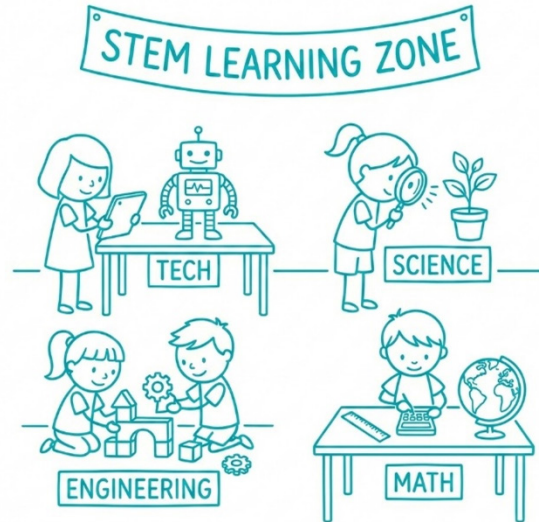
Safelnsights infrastructure to serve ECR communities

Considerations of use?

**Quest for fundamental
understanding?**

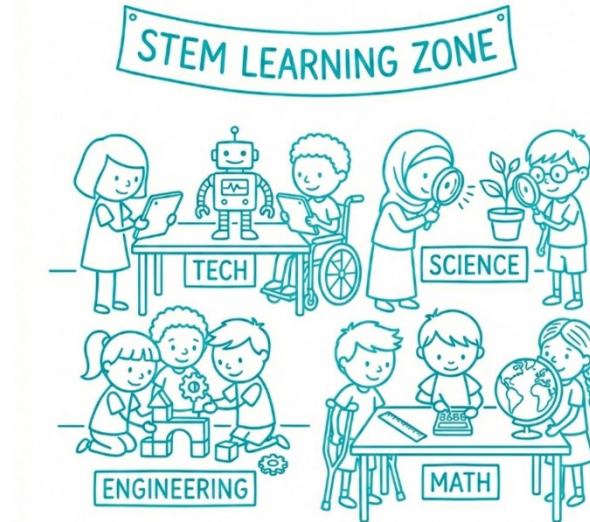
	No	Yes
Yes	Pure Basic Research	Use-inspired Research
No		Applied Research

ECR and SafeInsights have mutually aligned priorities



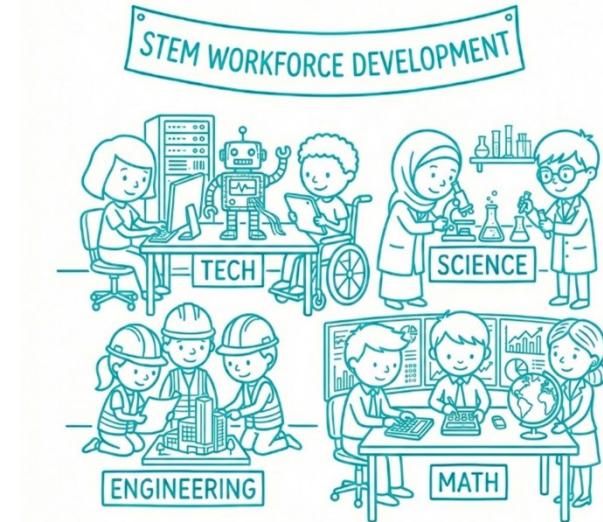
STEM learning and learning environments

The SafeInsights ecosystem includes data partners spanning K through grey levels representing different STEM learning contexts.



Broadening participation in STEM

SafeInsights aims to support emerging researchers across different disciplines with scaffolded supports. Complementing learning process with administrative data enables researchers to investigate how students farthest from opportunities learn and thrive.



STEM workforce development

SafeInsights aims to represent data partners spanning Education-to-workforce trajectories.

Where are we?

May 2024

- Awarded!
- Setting up team and subcontracts

March 2026

- Enclave build at various levels of maturity across partners
- Developing AI-backed coding and review support tool
- Significant progress on the logistical infrastructure

October 2026

- Operations ready
- More info coming in June at the ECR meeting!

Thank You.



Questions?

safeinsights.org

safeinsights.org

Federal, state, and other sources



Provider: National Center for Science and Engineering Statistics (NCSES) at NSF

- NCSES offers [a range of surveys](https://nces.nsf.gov/surveys) that contain information on the following topics (among others):
 - Government funding for science and education
 - Higher education research and development
 - Science and engineering workforce
 - » BA-recipients, PhD recipients, and the skilled technical workforce
 - » Includes training and educational experiences and academic/non-academic workforce experiences
 - Censuses of doctorate recipients and science/engineering graduate students and postdocs

<https://nces.nsf.gov/surveys>



Provider: National Center for Education Statistics (NCES) at the Department of Education

- Offers a [compendium of education-related surveys and administrative data collections](#). Topics include:
 - Characteristics of postsecondary institutions
 - BA recipients' subsequent educational and career pathways
 - Persistence and experiences in postsecondary education
 - Postsecondary transcripts
 - Transitions from high school to postsecondary education and the workforce
 - Use of financial aid
 - Graduate students
 - Postsecondary faculty
 - Cross-cohort comparisons available
 - Assessment of adult competencies in different domains, including numeracy

<https://nces.ed.gov/>



NCES continued: data sources for K-12 education

- NCES offers many such sources, including:
 - administrative data on public and private schools
 - national and international assessment programs
 - Geographic data on schools and communities
 - Teacher and principal workforces
 - Longitudinal studies tracking children in early childhood, the middle grades, and secondary education, with cross-cohort comparisons available
 - Household surveys of educational experiences for different family members including children



Provider: Bureau of Labor Statistics (BLS) at the Department of Labor

- [National Longitudinal Surveys](#), e.g., NLSY97 (teenagers followed from 1997 through 2021)
 - Education and training pathways; cognitive and non-cognitive measures; employment histories; earnings
- [O*NET](#)
 - Information on occupational outlooks, skills required, earnings in occupation
- Many, many reference sources on workers, employers, industries



Provider: Census Bureau at the Department of Commerce

- Population surveys
 - American Community Survey (annual cross-section)
- Labor force surveys
 - Current Population Survey (monthly labor force survey)
 - Survey of Income and Program Participation (longitudinal household panel)
- These surveys often don't have incredibly detailed educational and/or occupational information
- Census data can be very useful when your questions have geographic elements



Census Bureau, continued

- [Post-Secondary Employment Outcomes](#) (PSEO)
 - Tool with aggregate data
 - Includes earnings outcomes by major for participating colleges and states and different levels of institutions
 - Also features flows from major to industry and to state/region
 - Tool creates charts and graphs, and underlying aggregates are downloadable
 - Part of the [Longitudinal Employer-Household Dynamics](#) program



State Longitudinal Data Systems (SLDS)

- Data for people within a state that links individuals' records across agencies and over time.
- What is included varies by state
- Often includes data in some of these domains:
 - Apprenticeships
 - Early learning
 - K-12
 - Teacher workforce
 - Postsecondary and financial aid
 - Workforce (most commonly, Unemployment Insurance records)
 - Special populations and programs (foster care; juvenile rehabilitation; other program participation)



SLDS, continued

- Data come from a variety of sources including relevant state agencies, education institutions, Census Bureau, assessment companies, and NCES collections about schools and institutions
- Lots of variation across states:
 - The name of the system
 - Which data sources are included, and for how long historically
 - If and how external researchers can access (e.g., application process)
 - Costs to access



Student engagement/student experience surveys

- These surveys offer more fine-grained information on education experiences in postsecondary education (e.g., participation in high impact practices). They are typically not nationally representative but may have a large number of participating institutions.
 - [National Survey of Student Engagement](#) and its related studies
 - [Higher Education Research Institute](#) at UCLA has multiple studies
 - [Student Experience in the Research University](#), Undergraduate and Graduate versions
- These organizations often have faculty surveys as well
- Many offer different surveys depending on the institutional population (e.g., 4-year, 2-year) and/or level of student
- You may need to pay for data extracts



Considerations when using secondary data



Status of various data sources today

- [EssentialData.U.s](#) tracks confirmed data terminations and removed variables.
- Be aware of when data were collected to ensure the timeframe is appropriate for your investigation
 - NCSES indicates which collections are active and when the next collection is

Summary

The NSCG is a unique source for examining the relationship of degree field and occupation in addition to other characteristics of college-educated individuals, including work activities, salary, and demographic information.

Areas of Interest

Science and Engineering Workforce, STEM Education

Survey Administration

This survey was conducted by the Census Bureau in partnership with the National Center for Science and Engineering Statistics within the U.S. National Science Foundation.

Survey Details

Status	Active
Frequency	Biennial
Reference Period	The week of 1 February 2023
Next Release Date	19 January 2027

Report Downloads



Status of various data sources today

- NCES does not currently accept applications for restricted-use data access (consider partnerships with people who already have this access)

ResearchDataGov is moving to the new **Standard Application Process (SAP) Portal**. Starting **May 26 at 3:00 p.m. ET**, ResearchDataGov.org will enter a read-only state. You can still search the data catalog and view your dashboard, but you cannot start or edit applications. The new **SAP Portal** launches on **June 15**. On that date, you can resume all activities—including submitting and editing requests—on the new site. **ResearchDataGov.org will close permanently on June 16.**

Note: the Internal Revenue Service Statistics of Income did not have a 2025 Joint Statistical Research Program application window and is not accepting new proposals, with a re-evaluation of the program's status to occur at a date to be determined.

★ This action does not impact requests to access commingled Federal Tax Information through the Federal Statistical Research Data Centers.

Note: the National Center for Education Statistics is not accepting new proposals or reviewing proposals.

Note: the SAMHSA Center for Behavioral Health Statistics and Quality is accepting applications for the National Survey on Drug Use and Health.

About How To Apply FAQ Agency Info & Metrics Contact Login



ResearchDataGov

ResearchDataGov is a web portal for discovering and requesting access to restricted microdata from federal statistical agencies

Search for a keyword



[Researchdatagov.org](https://researchdatagov.org) is the site where users can request access to restricted microdata from federal statistical agencies



Data documentation

- Think carefully about what documentation is available
 - Can you find it? (Sometimes this is harder than it may seem!)
 - Does it include details on sampling? Data collection dates? Sample sizes? Content?
 - Can it answer questions about the data generation process that will inform your interpretation of results?
 - » Think about administrative collections and the complexities of who is contributing to them and where their data and information come from!
 - Is there enough information to understand different types of missing data that may occur, and how those are identified in the dataset?



Codebooks

- Hopefully, your codebooks and documentation contain these elements:
 - Some kind of instrument facsimile/specifications
 - » This is needed to understand the exact question, and the rules for who answered the question
 - How that question maps onto variables in the dataset (e.g., variable names)
 - Variable values
 - Value labels
 - Helpful: descriptive statistics for the variables, so you can see sample sizes and distributions



Missing data

- Be attentive to different kinds of missing data
 - e.g., why might data in a certain field be missing in an administrative collection? Is it the same reason for every case where the information is missing?
 - Where some participants intentionally skipped on a survey question?
- Does the codebook identify different types of missing data, for example, with different reserve codes?
- This is where understanding the data generation process is very helpful!

Missing data can be an issue in both primary and secondary data collections. In a primary data collection, you may be in a better position to understand its cause!



Privacy treatments

- Many datasets that include unit-level data are treated to avoid deductive disclosure
 - Deductive disclosure: being able to back out the identify of a participant in a deidentified dataset
- The smaller your focal population, the more the privacy treatments are likely to affect the point estimates
- Especially important to understand this if you are looking at:
 - Small geographic areas
 - Very small populations
- Review the documentation and recommendations provided with your data



Weighting and variance estimation

- Weights are used to indicate the relative contribution of an observation
 - E.g., a weight of 500 versus a weight of 100 or even a weight of 1
- In the simplest case, each observation contributes equally
 - This occurs in simple random samples (SRS) where every case has an equal chance of being included in the study
- In other cases, observations may contribute differently
 - This occurs in complex samples where cases have different probabilities of selection due to differential sampling of the population

In complex sampling situations, analyses based on weighted data are needed to produce findings that represent the target population



What do weights do?

- Adjust for differential selection probabilities
- Reduce bias associated with nonresponse by adjusting for differential response



(How) Should I use weights?

- If you are working with data from complex sampling, and you want to produce estimates that are representative of the target population, yes!
- Your dataset's documentation should explain the sampling procedures, the weights that were produced, and how to use them
 - Many types of documentation even include code for different software packages to help you implement the weighting
- Some datasets will have many different possible approaches depending on the data elements you include in your analysis.
 - The documentation should help you make these decisions



Complex samples and standard errors

- Usual standard error (SE) formula assumes a simple random sample (SRS)
- Software designed for SRS tends to underestimate standard errors in complex samples, because cases within sample groups tend to be similar to each other
- Special software/statistical approaches can adjust standard errors to take complex sampling into account. For example:
 - Replication techniques (“replicate weights”)
 - Taylor series linearization (“PSU,” “stratum”)



Data access, training, and support



Data access considerations

- Different levels of access to data are available
 - Public-use micro-data (unit-level data)
 - » Some variables and categories may be suppressed or aggregated
 - Publicly available aggregates
 - Publicly available through an online data tool (i.e., analyses on the microdata are available, but researchers cannot download the microdata)
 - Restricted-use data (unit-level data)
 - » Typically would include the full list of variables and detailed categories
 - » But still has privacy treatments
 - » Application process required (varies by agency) and accessed through a portal, typically
 - Data extracts available (i.e., a random sample but not the full dataset)



Downloading public-use data

- Datasets may be offered in a variety of formats (e.g., .csv, SAS, Stata, SPSS, etc.)
- In some cases, you will download a flat file and then download a program that will compile the dataset into the correct format
- You may receive the entire file, or you may use a tool to select cases and variables
- Dataset websites and documentation will help you determine how to go about this



Resources to find training and support

- [NCSES Resource Library](#)
 - Includes brief introductions to all surveys and links to additional resources
- [NCSES Knowledge Base](#)
 - More focused on data and data tools
- [NCES Distance Learning Dataset Training](#)
 - Covers content, design, and analysis considerations for longitudinal studies, assessments, and administrative data collections
- [BLS help and tutorials page](#)



Your next steps

- Start exploring the different data sources and datasets
- Consult data archives such as the [PEERS Data Hub](#) and [ICPSR](#)
- Read through some documentation to better understand how information may be presented
- Download some data and start investigating!
- Phone a friend to collaborate and/or get advice from others who have used the data before



Resource sharing and questions



Helpful links

- [All NCSES surveys](#)
- [NCES longitudinal studies program](#)
 - Includes early childhood, high school, and postsecondary studies
- [NCES homepage](#), which links to many other types of data they collect
- [Sign up for NCES Newsflash](#) – email blast when reports and new data are released
- [Get email updates from NCSES](#)
- DataLab – [NCES online data tool](#)
- NCSES [online data tools](#)

Get e-mail updates from NCSES

NCSES is an official statistical agency. Subscribe below to receive our latest news and announcements.

Subscribe →





Thank you!

ECR Hub
ecrhub.org

SafeInsights
safeinsights.org