Request for Solutions Informational Webinar

Topics:

- Creation of Synthetic Data for the Survey of Earned Doctorates and Development and Use of Verification Metrics
- Utilizing Privacy Preserving Record Linkage to Link Data from Two Federal Statistical Agencies
- Utilizing Privacy Preserving Record Linkage with Parent Agency Data and Statistical Agency Data to Inform Programs and Policies

Dr. Heather Madray

National Center for Science and Engineering Statistics (NCSES), Program Director for Data Access, Confidentiality, and Quality Assessment (DACQA)

Lisa B. Mirel NCSES, Statistical Advisor







Housekeeping Items:

- All attendees are on mute and will not be able to unmute themselves.
- Please use the "chat" function for technical difficulties only.
- Place all questions in the Q&A Box.
- Please check your audio settings if you are having difficulties hearing us.



Background

Dr. Heather Madray NCSES, Program Director for DACQA







Placed within NSF's SBE Directorate, NCSES is one of 13 principal federal statistical agencies







NCSES's mission is to serve as a federal clearinghouse for objective data that provide key insights into the S&E enterprise



Science & Engineering Workforce

Research & Development



STEM Education



Government Funding for S&E



Innovation & Global Competitiveness





The idea of a transformational NSDS to coordinate data linking, secure access, and support innovation everywhere has a distinguished lineage





America's DataHub provides experiences to inform the development of the NSDS Demonstration (NSDS-D) Project





Scope of the NSDS-D

e	To pilot potential shared services, technologies, and techniques that might be utilized within a potential
ie -	NSDS.

Focus is on novel research collaborations, data linkage methodologies, and privacy preserving technologies and techniques.

Exploring innovation is key to the NSDS-D.

The NSDS-D is **NOT** a data warehouse.

Data will not be collected and housed under the demonstration project.



Requests for Solutions Topics







Creation of Synthetic Data for the Survey of Earned Doctorates and Development and Use of Verification Metrics (SEDSyn)

Dr. Heather Madray NCSES, Program Director for DACQA



SEDSyn: Objective

The objective of this project is to produce a synthetic data file for public use to support a tiered access model, explore the use of synthetic data for evidencebuilding, and test the use of verification metrics in validating estimates produced from synthetic data.

A synthetic data set does not contain the exact records of the original dataset, but instead retains the statistical properties of the original dataset, preserving information useful to researchers and their queries.



SEDSyn: Background

As demand for access to confidential federal data assets increases alongside novel analytical approaches, privacy protections must be in place to ensure the protection of privacy and the confidentiality of the data. Use of synthetic data can reduce disclosure risk while allowing data users to access microdata for research and other statistical purposes.

The Survey of Earned Doctorate (SED) is a census of all individuals receiving a research doctorate from U.S. academic institutions in a 12-month period from July 1st to June 30th of the data collection year. These data are highly valuable in studying educational trends and differences by demographic groups including studies of equity. Currently, a restricted-use version is available for researchers who are approved for use, enter into a licensing agreement with NCSES, take annual training, and use the data in a secure virtual environment.

The proposed availability of a synthetic dataset would provide an additional tiered access option that would allow researchers to use SED microdata without an application or restricted use data license. This option could prove valuable to researchers in making maximum use of these data while enabling the government to ensure privacy.



SEDSyn: Steps for Completion

The following steps will be completed to explore the development of synthetic data:

- 1. Initial assessment to determine what variables for a synthetic dataset would be most useful for current and potential policymakers, researchers, data users and other stakeholders.
- 2. Open-source packages, such as R packages, would be utilized to create a synthetic version of the SED to produce a public-use microdata file. All contract staff accessing data will need to become CIPSEA agents and take annual data security training.
- 3. Once produced, the dataset would be evaluated to assess quality and disclosure risk. Use cases for this synthetic data would be identified to assist in determining how the synthetic data could be optimally used.
- 4. To assess the alignment of the synthetic data with the restricted data and to assess the quality of estimates produced from the synthetic data, verification metrics would be developed so that policymakers, researchers, data users and other stakeholders would be able to request them on an as needed basis.
- 5. A plan for dissemination of the synthetic data, verification metrics, and public messaging regarding this new public-use data product would be developed in conjunction with stakeholder outreach and feedback.



SEDSyn: Information Gaps

This project will identify key components necessary to inform a future/potential NSDS including:

- If a synthetic dataset can be produced on a demographic dataset using open-source packages.
- How verification metrics can be used to inform the quality of synthetic data in evidence-building.
- A path to produce verification metrics for an agency's restricted data and transfer those metrics to a centralized shared server. Use cases for a synthetic version of a high-value demographic dataset.
- A model for dissemination of a synthetic dataset in a tiered access model and messaging surrounding potential uses.



SEDSyn: Key Evidence Building Considerations

Key focus questions (address one or more) to assess innovation in the following areas: data acquisition, data security, data linking, privacy, and engagement:

- Which novel techniques for data, privacy, and confidentiality protections can be used?
- Are the resulting data and models fit to inform policy discussions and to make data available more equitability?
- What types of collaboration and stakeholder engagement are needed to help inform these questions, data, and analysis?
- What mechanisms are needed to access the resulting data that uphold privacy requirements?



SEDSyn: Deliverables

At a minimum, offerors will provide the following if selected for an award. Additional deliverables may be required.

- Monthly report, using NCSES template, to document progress.
- Stakeholder outreach summary.
- All code (clearly documented), documentation of synthetic data methodology, documentation of data quality assessment, and any other data/documentation created under this award.
- Documentation of verification metrics and alignment with true estimates.
- A report describing the lessons learned through this project including but not limited to the creation of synthetic data and whether the resulting data and models fit are able to inform policy discussions. In addition, the report should describe how this approach could inform a tiered access model and contribute to a potential National Secure Data Service.



Utilizing Privacy Preserving Record Linkage to Link Data from Two Federal Statistical Agencies (PPRL1)

Lisa B. Mirel NCSES, Statistical Advisor



PPRL1: Objectives

- 1) To develop a data sharing agreement between two federal statistical agencies.
- 2) To utilize a privacy preserving record linkage (PPRL) tool to link NCHS National Hospital Care Survey (NHCS) and the NCSES Survey of Earned Doctorate (SED) through a trusted third party hosted by NCSES. This linkage would serve as an example of two federal agencies utilizing PPRL for interagency data sharing and linking which will inform future interagency initiatives.
- 3) To demonstrate the ability to analyze the linked data in a secure environment. The resulting linked file will be used to provide validation statistics back to NCHS to provide additional insight on the validity of the education imputation methodology.



PPRL1: Background

The Advisory Committee on Data for Evidence Building (ACDEB) was charged with making recommendations to the OMB Director on how to promote the use of federal data for evidence building. Those recommendations include identifying mechanisms for streamlining data-sharing agreements across federal agencies, promoting the use of privacy-preserving technologies in a tiered access framework, and promoting the use of privacy preserving technologies that can support working with data in situ.

To build on these recommendations, this project will serve as a proof of concept to develop a data sharing agreement between two federal statistical agencies that have not previously developed data sharing relationships, deploy a PPRL tool to link data from two federal statistical agencies and utilize a secure environment to analyze the resulting linked data file.



PPRL1: Information Gaps

This project will identify key components necessary to inform a future/potential NSDS, including:

- What is needed to create a standardized data sharing agreement between two federal statistical agencies, one of which would serve as a trusted agent who would host the compute environment to conduct the linkage.
- What is needed to conduct PPRL to link these sources without ever exchanging direct personally identifiable information.
- What is needed to build an infrastructure to inform linkages across the federal government, using the development of agreements, utilization of a central, shared compute environment, and deployment of PPRL as a model to improve the availability, quality, accessibility, and interoperability of data sharing and linking.
- What is needed to assess the feasibility of using linked data to inform/validate imputation models that will support evidence-based policymaking.



PPRL1: Key Evidence Building Considerations

What are key challenges with locating, acquiring, accessing, linking, and using disparate data and information?

- What are key challenges with locating, acquiring, accessing, linking, and using disparate data and information?
- What are the lessons learned to using the presumption of accessibility in the Evidence Act?
- Which novel techniques for data, privacy, and confidentiality protections can be used?
- What models and estimation methods are best suited to fill information gaps?
- Are the resulting data and models fit to inform policy discussions and to make data available more equitability?
- What types of collaboration and stakeholder engagement are needed to help inform these questions, data, and analysis?
- What mechanisms are needed to access the resulting data that uphold privacy requirements?



PPRL1: Deliverables

- Selection of a PPRL tool that meets certain minimum requirements, including but not limited to successful
 past performances with documented high precision and recall. The tool should demonstrate performance
 linking with identifiers such as partial social security number, name, and address.
- All code (clearly documented), linked data file, documentation of linkage process, documentation of linked data file (including codebook), and any other data/documentation created under this award.
- A report describing the lessons learned through this project including but not limited to the selection of the PPRL tool, and the key challenges/opportunities with locating, acquiring, accessing, and linking data using PPRL.



Utilizing Privacy Preserving Record Linkage with Parent Agency Data and Statistical Agency Data to Inform Programs and Policies (PPRL2)



PPRL2: Objectives

- 1) Develop a data sharing agreement between a federal statistical agency and its parent agency.
- 2) Utilize privacy preserving record linkage methods to connect SED and principal investigator (PI) data for awarded proposals in a secure server environment (e.g., a trusted third party managed by NCSES). This linkage would serve as an example of utilizing PPRL for data sharing and linking which will guide future interagency initiatives.
- 3) Demonstrate the ability to analyze linked data in a secure server environment. Once the linkage occurs the data will be analyzed to address key research questions and support the NSF learning agenda and other internal NSF stakeholder questions.



PPRL2: Background

Throughout the federal government there are many initiatives to link data for evidence building. One way to support these initiatives is to build the infrastructure necessary for cross agency linkages, including developing data sharing agreements, conducting linkages while protecting privacy, and utilizing secure environments for analyzing the resulting linked datafile(s).

To support these initiatives this project will serve as a research demonstration project for the NSDS to develop a data sharing agreement between a federal statistical agency and its parent agency, utilize a privacy protecting record linkage tool to link two disparate sources, and create an analytic dataset that can be used to answer questions that could not be answered with either source alone.

This project will include the exploration of different data sharing agreement options to identify templates for future data sharing opportunities. In order to conduct the linkage, a PPRL tool will be selected.

The two data sources to be linked using a PPRL tool are the NCSES SED and the PI data for awarded proposals maintained by NSF in its data systems. Once the data are linked using a PPRL tool and stripped of the encrypted tokens, they will be available for analysis in a secure environment.



PPRL2: Background, cont.

This work can also inform many of the objectives of the <u>NSF learning agenda</u>, such as:

- How could NSF leverage tools at its disposal— policies, strategies, programs, and so on—to increase the participation of these (most extremely underrepresented intersectional) groups in the STEM enterprise? Answers to these questions will help NSF identify best practices and align programs and policies toward closing gaps in participation in the STEM enterprise.
- How the work will facilitate analyses of data through a difference-in-differences approach (to measure differences in measures, such as proposals submitted by gender before and after the pandemic) and the specification of regression models as part of an interrupted time-series (ITS) design to determine changes that might be attributed to COVID— by modeling (and comparing) the expected pre-COVID and observed since-COVID trends, controlling for relevant factors.
- Findings will help NSF leadership and staff consider strategies for improving the efficacy and equity of the merit review process.



PPRL2: Information Gaps

This project will identify key components necessary to inform a future/potential NSDS including:

- The steps needed to develop a data sharing agreement between a federal statistical agency and its parent agency.
- What is needed to conduct PPRL to link these sources without ever exchanging direct PII.
- What is needed to build an infrastructure to inform linkages across the federal government, using the development of agreements, utilization secure server environment, and deployment of PPRL as a model to improve the availability, quality, accessibility, and interoperability of data sharing and linking.
- What is needed to assess the feasibility of analyzing linked data in a secure environment to support evidence-based policymaking.



PPRL2: Key Evidence Building Considerations

Key focus questions to assess innovation in the following areas: data acquisition, data security, data linking, privacy, and engagement:

- What are key challenges with locating, acquiring, accessing, linking, and using disparate data and information?
- What are the lessons learned to using the presumption of accessibility in the Evidence Act?
- Which novel techniques for data, privacy, and confidentiality protections can be used?
- What linked data models and estimation methods are best suited to fill information gaps?
- Are the resulting linked data and models suitable to inform policy discussions and to make data available more equitability?
- What types of collaboration and stakeholder engagement are needed to help inform these questions, data, and analysis?
- What mechanisms are needed to access the resulting data that uphold privacy requirements?



PPRL2: Deliverables

At a minimum, offerors will provide the following if selected for an award. Additional deliverables may be required.

- Develop a data sharing agreement between a federal statistical agency and its parent agency, including lessons learned to inform future efforts.
- Selection of a PPRL tool that meets certain minimum requirements, including but not limited to successful past
 performances with documented high precision and recall. The tool should demonstrate performance linking with
 identifiers such as name, email, and address.
- Creation of two submission files that have been standardized for linkage
- All code (clearly documented), linked data file, documentation of linkage process, documentation of linked data file (including codebook), and any other data/documentation created under this award.
- Report summarizing findings from statistical analysis using the linked SED-PI data
- Final report summarizing all stages of the project and offering lessons learned for a potential future NSDS.



Questions?







.

Request for Solutions (RFS) Requirements

Rebecca Harmon Senior Contracts Manager Advanced Technology International (ATI), ADC Consortium Management Firm (CMF)

americasdatahub.org/opportunities

The official source of information regarding the solicitations is included in the Request for Solutions posted on the ADC website. If you act on information from other sources, it is at your risk.







RFS Summary

Project Topics

- Each proposals must address ONE of the specific topic areas.
 - ➢ If you intend to respond to all three, each RFS must be submitted separately.
- See Attachment 1 of each RFS for full topic description.

Anticipated Funding

- **SEDSyn:** One or more awards estimated at \$800,000
- PPRL 1: One or more awards estimated at \$750,000.
- PPRL 2: One or more awards estimated at \$750,000.

Period of Performance

• Limited to 24 months for each project



One Step Process

- Offerors will submit a detailed technical and cost proposal for award evaluation.
- You do not need to be an ADC member to respond. However, if you're selected for award, you must join ADC (if not already a member).



Full Proposal Submission

- RFS Attachment 2 includes format
 - Volume 1: Technical Proposal
 - Limited to **8 pages plus cover page.**
 - Volume 2: Cost Proposal
 - No page limit.
 - Submit in Word format.
 - Submission forms:
 - SEDSyn: https://atisc.formstack.com/forms/adc_sedsyn_rfs
 - PPRL 1 : https://atisc.formstack.com/forms/adc_pprl1_rfs
 - PPRL 2: https://atisc.formstack.com/forms/adc_pprl2_rfs



Full Proposal Cover Page

- Working title of the proposed project.
- Names, phone numbers, mailing, and e-mail addresses for the principal technical and contractual points of contact (person or persons authorized to negotiate on the behalf of the offeror and who can contractually obligate the offeror organization).
- Unique Entity ID (formally DUNS number) of the submitting organization (if available).
- Project partners, if any.
- Date of submission.
- Proprietary data restrictions, if any.



Volume 1: Technical Proposal Content

Executive Summary

- Summary Statement: Provide a succinct statement of the aim of the project and proposed approach. In most cases, the summary statement will be no longer than a paragraph.
- **Context:** Briefly describe the current state of information and/or research in the area.
- Proposed Approach: Describe the proposed approach. Offerors shall explain how the proposed approach will meet the objectives outlined in Attachment 1; result in or lead to a replicable framework that can be used to address similar issues; and inform other strategic priorities like the National Secure Data Service.


Statement of Work

- Work Scope: Describe the work to be accomplished as part of the project, organized as it is expected to be performed. Separate the work effort into major tasks and subtasks as numbered paragraphs or in a table.
- **Deliverables:** All project deliverables should be clearly listed and described.
- Future Phases: Proposals may include a discussion of optional, future phases of work. The original phase or work shall in no way depend on work described under future phases to meet the program criteria.
- There shall be no company-sensitive or proprietary data included in the Statement of Work.



• Capabilities and Experience

- List all team members proposed for the project and indicate if they are a non-traditional entity as defined in the RFS.
- List all key personnel, including those from outside the offeror's organization. Organize the team by organization name and briefly describe each person's roles and responsibilities on the project.
- Identify and describe the capabilities and experience of key personnel and organizations as these elements relate to the proposed project.
- Descriptions of experience should serve to demonstrate the key personnel's ability to successfully conduct the proposed research or project, including access to critical resources for the project.
- Designate any supervisory relationships and who will be the main point of contact for regular check-ins with the NCSES methodological team during the project.
- Provide condensed resumes (2-page maximum) for all key personnel on the project.



Capabilities and Experience

- In additional to key personnel, designate any graduate students or postdoctoral fellows funded by the proposed research. If named, provide no more than a half-page biographical sketch of their background and research interests. The biographical sketch should be included as part of the resume appendix and does not count against the page limit.
- Describe any unique capabilities that the offeror team possesses that may reduce project risk, reduce project duration, and/or improve project financial performance.
- Address any potential conflicts of interest and any proposed mitigation. In addition, complete and submit Exhibit 1 – Organizational Conflicts of Interest Certificate.



Intellectual Property Rights

- Describe any limitations on any intellectual property (patents, inventions, trade secrets, copyrights, or trademarks) that will impact the Offeror's performance of the contract or impact the Government's subsequent use of any deliverable under the contract.
- The Offeror must describe the intellectual property in sufficient detail to describe the limitations (data assertions of the Offeror or any subcontractor, potential patent licenses required by the Government, etc.), and to describe why or how the Government can accomplish the stated objectives of this RFS with the limitations described or proposed by the Offeror.



Volume 2: Cost Proposal Content

• The Cost Proposal shall contain the following information:

- Agreement Type: It is anticipated the proposed efforts will be funded as firm-fixed-price. However, offerors
 may recommend an alternate approach (e.g., cost, cost-plus-fixed-fee, etc.) and include the rationale for
 their use. State either firm-fixed-price or provide rational for an alternative approach.
- **Cost Estimate:** The cost estimate shall be broken down for each year of the proposed work. The elements of cost and suggested level of detail are listed in the RFS. Contractor format for the cost estimate is acceptable.



Volume 2: Cost Proposal Content (Continued)

• Program Funds

- Labor Offeror only: Provide a description of each labor category or person with associated labor rate and hours.
- Travel Offeror only: Provide a list of the number of trips, number of days and travelers per trip, cost per trip, and the purpose of each trip.
- Team Members/Subcontractors/Consultants: Provide a list of each team member/subcontractor/consultant and their associated subcontract funding.
- Material/Equipment Offeror only: Provide a list all items and provide justification and basis of cost for each (i.e., catalog pricing, vendor quote, previous purchase, etc.).
- Other Direct Costs Offeror Only: Provide a list all items and provide justification and basis of cost for each (i.e., catalog pricing, vendor quote, previous purchase, etc.).
- Indirect Costs Offeror Only: Provide a breakout of all indirect costs and indicate whether rates are Government approved, citing approval date and federal agency providing approval. If rates are not approved, provide an explanation on how your organization's proposed rates are appropriate for pricing.
- Profit/Fee: Indicate any profit/fee.



Vol 2: Cost Proposal Content (Continued)

- Cost Share (not required)
 - Cash (labor, travel, etc.): As applicable, provide a description of each labor category, details on travel, list of materials, etc.
 - In Kind (use of equipment, space/ buildings, intellectual property): Must provide basis of cost.



Full Proposal Submission Form

Offeror Information

. .

Offeror Organization *

Offeror Address

Offeror City

Offeror State

Offeror Zip Code



Full Proposal Submission Form (Continued)

Are you currently a "Non-traditional Entity"?*



No

"Non-traditional entity" means an entity (construed in its broadest sense to include qualified large and small businesses, universities, non-profits, philanthropic organizations, partnerships, joint ventures, and other entity forms) that is not currently performing and has not performed, for at least the three-year period preceding the solicitation of sources by NSF for the procurement or arrangement, under any NSF procurement contract or NSF instrument of financial assistance.



Full Proposal Submission Form (Continued)

POC Information

Technical	POC	Name *	

Į		
First Name	Last Name	

Job Title*

Technical POC Email*

Technical POC Phone*

Is the Contracts POC the same as the Technical POC?



Full Proposal Submission Form (Continued)

Submission

Proposal Title*

Proposal Submittal

O I prefer to upload the proposal to this form (attachments will be unencrypted)

I prefer to send the proposal via encrypted email to adc-contracts@ati.org (may be done before or after submitting this form)

Volume I: Technical Proposal Upload

Remove File No File Chosen Choose File File names must not contain spaces or special characters

Volume II: Cost Proposal Upload



File names must not contain spaces or special characters

Additional Comments



Full Proposal Evaluation Criteria

Technical

- Approach the degree to which the proposed project:
 - (i) meets the objectives outlined in (RFS) Attachment 1;
 - (ii) will result in or lead to a replicable framework that can be used to address similar issues; and
 - (iii) demonstrates innovativeness
 - o (iv) informs other strategic priorities like the National Secure Data Service.
- Teaming The degree to which the proposed project includes a diverse team of qualified performers to include use of non-traditional entities. (See RFS for nontraditional definition)

• Cost

- The CMF will perform an analysis and will provide the results to the Government. This
 effort may entail the CMF requesting additional information from the Offeror.
- The Government will determine whether the Offeror's total evaluated cost/price is fair and reasonable.

The criteria are listed in order of relative importance.



Base Agreement



If selected for award, the Base Agreement must be signed at that time.



Any Project Agreements will be awarded under the Base Agreement.



Timeline

	Date (Estimate)
Request for Solutions Release	June 21, 2023
Webinar	June 26, 2023
Proposal Deadline	July 12, 2023
Offeror Notifications	Late July 2023
Award Projects	August 2023

Any deadline updates will be communicated via email.



ADC Member Teaming Resources

- ADC Member Profile Database*
 - Searchable by member demographics and capabilities, includes POC info for each member.
- Member Forum*
 - Post under designated forum if you are interested in teaming or starting a discussion with other members.
- Need a teaming partner outside of ADC or other resources?
 - Email ati@govmates.com with who you are looking for.
 - More information is available at <u>govmates.com/ati</u>.

*Requires access to Members Only website. Request access here.

Not a member, but want to access these resources? Join today — it's free!



Communications

RFS Questions: <u>ADC-Contracts@ati.org</u>



General/Membership Questions: adc@ati.org



Visit the ADC Website

Stay in the loop with ADC events, solicitations, project awards, news, members, and more!

www.americasdatahub.org





Questions?







.