# Creation of Synthetic Data for the Annual Business Survey (ABS) and Use of Verification Metrics
## and
# Models for a Data Concierge Service for a National Secure Data Service

**Requests for Solutions Webinar**

**Dr. Heather Madray**
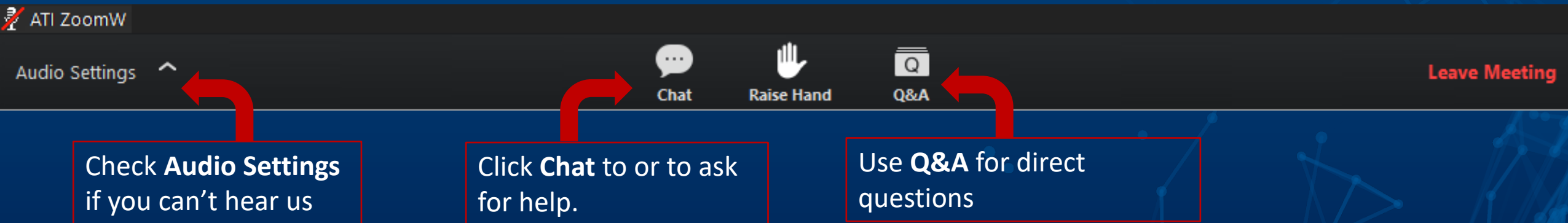NCSES, Program Director for Data Access, Confidentiality, and Quality Assessment (DACQA)

# 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

ATI ZoomW

Audio Settings    Chat    Raise Hand    Q&A    Leave Meeting

Check **Audio Settings** if you can't hear us

Click **Chat** to or to ask for help.

Use **Q&A** for direct questions

AMERICA'S DATAHUB CONSORTIUM

NSF | NCSES

# Background

# Placed within NSF's SBE Directorate, NCSES is one of 13 principal federal statistical agencies
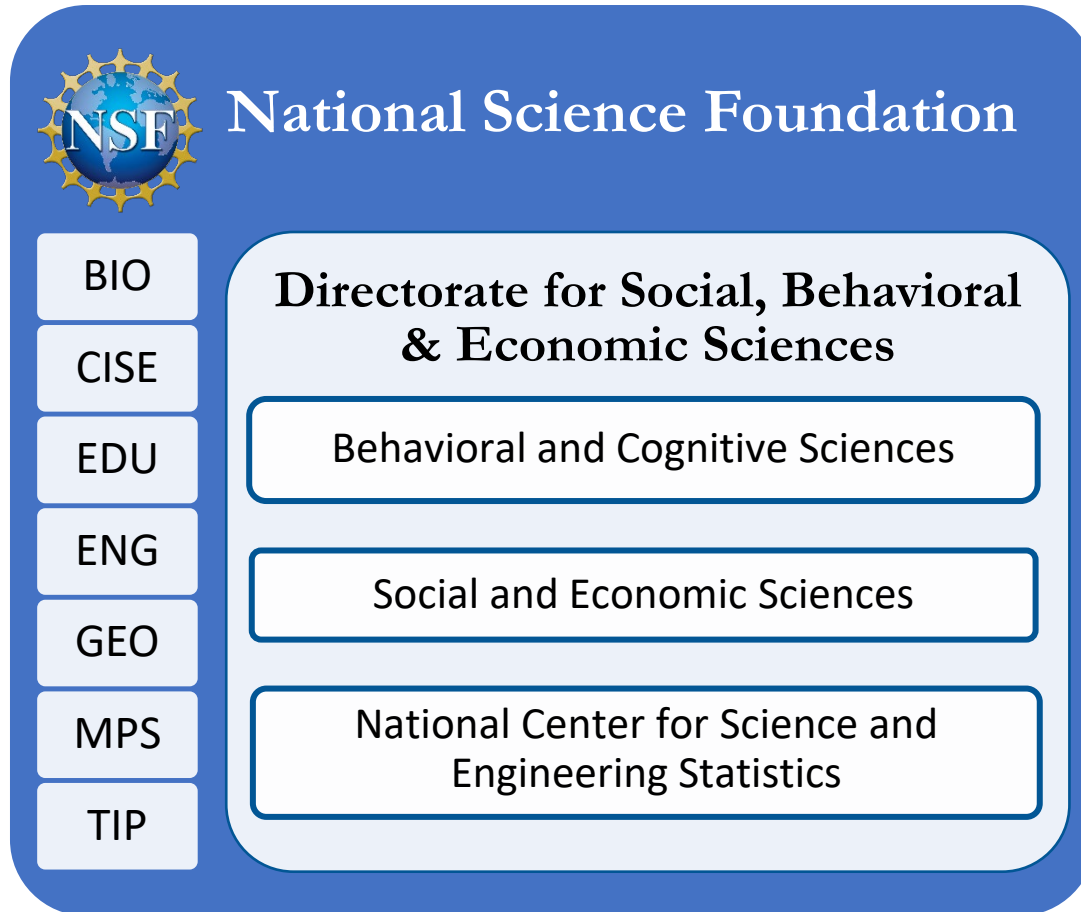
# NCSES's mission is to serve as a federal clearinghouse for objective data that provide key insights into the S&E enterprise

**Science & Engineering Workforce**

**STEM Education**

**Innovation & Global Competitiveness**
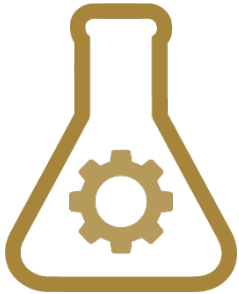
**Research & Development**

**Government Funding for S&E**

**Higher Ed R&D**

AMERICA'S DATAHUB
CONSORTIUM

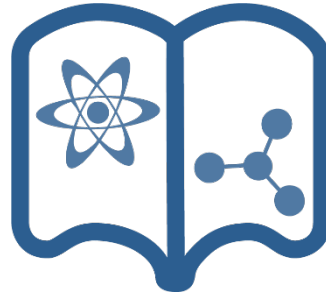# The idea of a transformational NSDS to coordinate data linking, secure access, and support innovation everywhere has a distinguished lineage

The pathway to the
## NATIONAL SECURE DATA SERVICE

**2016**      **2018**      **2020**      **2021**      **2022**
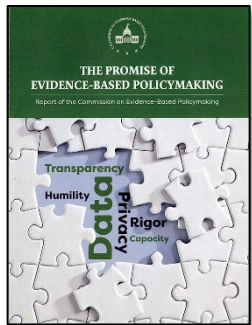
**Commission on Evidence-Based Policymaking**

Foundations for Evidence-Based Policymaking Act of 2018

ACDEB
Advisory Committee on Data for Evidence Building

AMERICA'S DATAHUB
CONSORTIUM

The CHIPS and Science Act

**National Secure Data Service Demonstration Project**
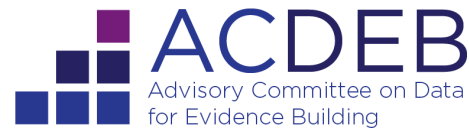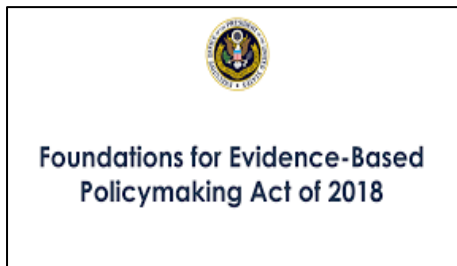
AMERICA'S DATAHUB
CONSORTIUM

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

**NSF | NCSES | ICSP**

**NSDS DEMONSTRATION PROJECT**

Broadened engagement and opportunities for innovation

Validate shared-service operational model

NSDS-D Website

Test NSDS features/ functions

**CHIPS Act:** Establish a demo project to develop, refine, and test models to inform the full implementation for a government-wide data linkage and access infrastructure for statistical activities conducted for statistical purposes.

**AMERICA'S DATAHUB CONSORTIUM**

Flexible acquisition process & mechanisms for delivering services

Broadened engagement and opportunities for innovation

Agility to address emerging research topics

**NSF Strategic Objective 4.1: Strengthen at Speed and Scale**
Pursue innovative strategies to strengthen and expand the agency's capacity and capabilities.

ADC established August 2021

CHIPS Act authorizes 5-yr NSDS Demo

Report due to Congress

End of demo

| FY21 | ▲ FY22 | ▲ | FY23 | FY24 | ▲ FY25 | FY26 | FY27 ▲ | FY28 + |

# Scope of the NSDS-D

To pilot potential shared services, technologies, and techniques that might be utilized within a potential 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.*

**AMERICA'S DATAHUB**
CONSORTIUM

# Creation of Synthetic Data for the Annual Business Survey (ABS) and Use of Verification Metrics

**AMERICA'S DATAHUB**
CONSORTIUM

# Creation of Synthetic Data for the Annual Business Survey (ABS) and Use of Verification Metrics – Objective

To test and compare methods for creating synthetic data to support a tiered access model; explore the use of synthetic data for evidence-building; and test the use of verification metrics in validating estimates produced from synthetic data.

AMERICA'S DATAHUB
CONSORTIUM

# Creation of Synthetic Data for the Annual Business Survey (ABS) and Use of Verification Metrics

- The Annual Business Survey (ABS) is an annual survey conducted by the U.S. Census Bureau on behalf of multiple sponsors, one of whom is the National Center for Science and Engineering Statistics. The ABS includes, but is not limited to, data on business owners, including demographic information, research and development (R&D), innovation, and technology.

- One objective in disseminating ABS data is the production of a public use microdata file that could be used by researchers, policy makers, and public-data users.

- To address disclosure concerns, a synthetic datafile option is being explored.

- This project would build on this work to inform the National Secure Data Service Demonstration Project (NSDS-D) and support tiered access through the development of a publicly available dataset that could be used without barriers involved in accessing the restricted-use data.

AMERICA'S DATAHUB
CONSORTIUM

# Creation of Synthetic Data for the Annual Business Survey (ABS) and Use of Verification Metrics – Information Gaps

## This project will identify:

- What methods can be used to create synthetic datasets, focusing on ease of creation, data quality, and cost.
- 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 data and transfer those metrics to a centralized shared server.
- Use cases for a synthetic version of a high-value economic dataset.
- A potential path for collaborative opportunities for federal agencies in creation and assessment of synthetic datasets.

AMERICA'S DATAHUB
CONSORTIUM

# Creation of Synthetic Data for the Annual Business Survey (ABS) and Use of Verification Metrics – 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:

- What are the lessons learned to using the presumption of accessibility in the Evidence Act?
- Can synthetic data be used to support a tiered access model?
- 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?

AMERICA'S DATAHUB
C O N S O R T I U M

# Creation of Synthetic Data for the Annual Business Survey (ABS) and Use of Verification Metrics - Background

- Use of synthetic data can reduce disclosure risk while allowing data users to access microdata for research and other statistical purposes.

- A synthetic dataset 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.

- The anonymity of the original dataset is not compromised since synthetic records do not correspond to real ones.

- The production and use of synthetic data for the ABS is being explored to increase the utility of the data while ensuring strong privacy protections.

AMERICA'S DATAHUB
CONSORTIUM

# Creation of Synthetic Data for the Annual Business Survey (ABS) and Use of Verification Metrics – Background, *cont.*

- Currently, a restricted-use version of the ABS microdata is available for researchers who are approved for use by the Census Bureau, the Internal Revenue Service, and NCSES.
    - In addition to the initial approvals, researchers must obtain Special Sworn Status, take annual data stewardship training, and use the data in a Federal Statistical Research Data Center (FSRDC).
    - The administrative costs for government in managing projects and researchers that use this data is also significant, requiring proposal review, IT resources, and administrative resources to process Special Sworn Status applications for researchers.

- The availability of a synthetic dataset would provide a tiered access option that would allow researchers to use microdata without the requirements for secure access.

AMERICA'S DATAHUB
CONSORTIUM

# Creation of Synthetic Data for the Annual Business Survey (ABS) and Use of Verification Metrics

**This project will explore the following:**

1. A user workshop would be conducted to identify variables of interest for a synthetic public use file. This workshop will target both current data users given their experience and knowledge of the data as well as potential data users who could have interest in synthetic publicly available data. A report would be produced from this workshop that highlights user feedback and a potential initial list of variables or variable groupings that are of interest to ABS users.

2. Two methods of producing ABS synthetic data would be utilized to produce two separate synthetic ABS microdata datasets. One method, using proprietary CenSyn software, has been partially implemented and would require completion. The second method would use R packages to create a second synthetic version of the ABS public-use microdata file.

3. Once produced, the two datasets and the process for creating them would be compared to determine any differences in data quality and validity between the two products and the cost and ease of production given the two methods.

AMERICA'S DATAHUB
CONSORTIUM

# Creation of Synthetic Data for the Annual Business Survey (ABS) and Use of Verification Metrics, *cont.*

4. The Census Bureau Disclosure Review Board (DRB) in conjunction with the Internal Revenue Service, Statistics of Income Division (SOI), would review the chosen synthetic file for disclosure concerns.  Any publicly released dataset would require both Census Bureau and IRS-SOI approval prior to release which may require multiple iterations with some risk of rejection.

5. The public-use synthetic dataset(s) would be utilized for an evidence-building project in the NCSES-managed compute environment, with the use of verification metrics to test this method of verifying the quality of estimates produced from the synthetic data.  Use of the data in this environment will assist in testing a new secure compute environment.

AMERICA'S DATAHUB
CONSORTIUM

# Models for a Data Concierge Service for a National Secure Data Service

AMERICA'S DATAHUB
CONSORTIUM

# Models for a Data Concierge Service for a National Secure Data Service - Objective

The objective of this project is to explore models for a data concierge service for a potential, future National Secure Data Service (NSDS).

This concierge service would serve as a coordinator between data providers and data users, providing assistance to data users in answering general questions; identifying confidential data assets that meet their evidence-building needs; directing them to subject matter experts and more focused guidance when needed; and support in developing evidence-building proposals to apply for access to confidential data.

This project will explore the needs of data users who are investigating access to confidential data; the tools that would facilitate a data concierge service in assisting users; and the development of potential models for a data concierge service.

AMERICA'S DATAHUB
CONSORTIUM

# Models for a Data Concierge Service for a National Secure Data Service– Information Gaps, Considerations

## Information Gaps: This project will identify:

- What are potential models for a data concierge service that would support a National Secure Data Service.
- What types of service requests do federal agencies encounter and how can a concierge service best meet those needs.
- What tools are needed to support a data concierge service.

## Key Evidence Building Considerations

- Key focus questions (address one or more) to assess innovation in the following areas - user engagement and customer service:
  o What support do data users need to conduct evidence-building research?
  o How can a data concierge service be structured to meet the needs of evidence-building communities?

AMERICA'S DATAHUB
CONSORTIUM

# Models for a Data Concierge Service for a National Secure Data Service- Background

The year 2 report of the Advisory Committee on Data for Evidence-Building (ACDEB) provided several recommendations that focus on a data concierge service.

- Recommendation 3.2 focuses on the types of assistance likely sought by researchers, from general questions to data discovery to how to link multiple datasets.
- Recommendation 3.3 discusses potential tiers of assistance, such as the potential use of chatbots and other similar technologies to provide initial intake and answer simple questions.
- Recommendation 3.4 discusses using data concierges in the form of subject matter experts that can provide more in-depth assistance to identify confidential data needed for specific evidence-building needs, referrals to additional subject matter experts, proposal development, and providing overall technical expertise.

These three recommendations provide a high-level framework for the types of inquiries likely to be encountered in a data concierge service and a potential tiered approach to providing support.

AMERICA'S DATAHUB
CONSORTIUM

# Models for a Data Concierge Service for a National Secure Data Service – Background, *cont.*

- Data concierges may be required to provide support for:
  - Determining whether confidential data access is required and if so, what data sources are needed.
  - Determining the types of analysis required and how best to meet that need.
  - Providing guidance on data linkage.
  - Assistance with resource constraints, assisting the user in finding potential sources of funding or partnering.
  - Assistance in preparing a proposal may be needed.
  - Referrals to subject matter experts and agency officials, as needed.

- To ensure that data concierges can provide a high-level of evidence-building support, tools will be needed to assist data concierges with dataset availability, agency contacts, data linkage capabilities, modes of data access, and the range of options available to data users for their specific needs.  Examples of potential customer service

**AMERICA'S DATAHUB**
CONSORTIUM

# Models for a Data Concierge Service for a National Secure Data Service

This wide range of potential services requires a data concierge service model that takes into account the range of services needed, the types of service providers, such as human vs. chatbots, and the tools required by concierge providers to implement a comprehensive data concierge service.

Also, given the wide range of assistance, the limitations of a centralized data concierge service should be taken into account with a view towards what assistance can be provided by a concierge and what will need to be forwarded to agency contacts or other subject matter experts.

**AMERICA'S DATAHUB**
CONSORTIUM

# Models for a Data Concierge Service for a National Secure Data Service– Project Objectives

1. An environmental scan to determine how federal statistical agencies are currently providing customer service to researchers and data users and what types of support needs they encounter. This scan should include the identified challenges in providing support to users, not just traditional data users, such as within academic communities, but also non-traditional users such as those from non-profit organizations, state and local governments, and economic development organizations.. This scan will also include any existing data concierge services to inform potential models. This can include services outside of government.

2. Outreach to organizations of state data user networks, minority serving institutions, non-profit think tanks and research entities, and professional associations to identify the types of services and support needed by these users.

AMERICA'S DATAHUB
CONSORTIUM

# Models for a Data Concierge Service for a National Secure Data Service– Project Objectives, *cont.*

3.  Development of two or more models for a data concierge service, taking into account services available currently, data user needs, and the requirements for a shared services model, including anticipated cost and personnel to manage the service.  These models should include implementation guidelines for each including a user feedback mechanism to adjust services as user needs are further identified or change.

4.  Recommendations for tools for a data concierge service.  These tools could include decision-trees, contact lists, technical documentation on data and data linkage, and other supporting documentation or online tools to assist both users and data concierges in serving the data user community.

AMERICA'S DATAHUB
C O N S O R T I U M

# Models for a Data Concierge Service for a National Secure Data Service– Deliverables

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

- Monthly status reports on progress towards project objectives.
- Report detailing environmental scan and user outreach efforts and outcomes.
- Two or more models for a data concierge service, with implementation plans for each.
- Recommendations for tools that can be developed and utilized for a data concierge service.

**AMERICA'S DATAHUB**
C O N S O R T I U M

# Questions?

# Request for Solutions (RFS) Requirements

**Ms. 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.*

AMERICA'S DATAHUB
CONSORTIUM

NSF | NCSES

# RFS Summary

## Project Topics

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

## Anticipated Funding

- **ABS Synthetic Data:** One or more awards anticipated, total estimated at $1,000,000.
- **Data Concierge:** One or more awards anticipated, total estimated at $500,000.

## Period of Performance

- **ABS Synthetic Data:** Limited to thirty (30) months.
- **Data Concierge:** Limited to eighteen (18) months.

AMERICA'S DATAHUB
CONSORTIUM

# 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).

**AMERICA'S DATAHUB**
**C O N S O R T I U M**

# 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:
    - ABS Synthetic Data: https://atisc.formstack.com/forms/adc_abssyn_rfs
    - Data Concierge: https://atisc.formstack.com/forms/adc_dcs_rfs

AMERICA'S DATAHUB
C O N S O R T I U M

# 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.

AMERICA'S DATAHUB
CONSORTIUM

# 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.

**AMERICA'S DATAHUB**
CONSORTIUM

# Volume 1: Technical Proposal Content *(Continued)*

- ## 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.***

**AMERICA'S DATAHUB**
CONSORTIUM

# Volume 1: Technical Proposal Content *(Continued)*

- **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.

**AMERICA'S DATAHUB**
C O N S O R T I U M

# Volume 1: Technical Proposal Content *(Continued)*

- ## 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.

AMERICA'S DATAHUB
CONSORTIUM

# Volume 1: Technical Proposal Content *(Continued)*

- **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.

AMERICA'S DATAHUB
CONSORTIUM

# 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.

AMERICA'S DATAHUB
C O N S O R T I U M

# 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.

AMERICA'S DATAHUB
CONSORTIUM

# 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.

**AMERICA'S DATAHUB**
CONSORTIUM

# Full Proposal Submission Form

## Offeror Information

Offeror Organization *

[ ]

[ ]

Offeror Address

[ ]

Offeror City

[ ]

Offeror State

[ ]

Offeror Zip Code

AMERICA'S DATAHUB
CONSORTIUM

# Full Proposal Submission Form *(Continued)*

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

☐ Yes

☐ 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.

**AMERICA'S DATAHUB**
CONSORTIUM

# Full Proposal Submission Form *(Continued)*

## POC Information

Technical POC Name *

| | |
|---|---|
| First Name | Last Name |

Job Title *

Technical POC Email *

Technical POC Phone *

Is the Contracts POC the same as the Technical POC?

AMERICA'S DATAHUB
CONSORTIUM

# Full Proposal Submission Form *(Continued)*

Submission

Proposal Title *

[                                                                    ]

Proposal Submittal

○ 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

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

Volume II: Cost Proposal Upload

[Choose File] [Remove File] No File Chosen
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
    - (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 non-traditional 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.*

AMERICA'S DATAHUB
CONSORTIUM

# ADC Member Teaming Resources

- ADC Member Profile Database*

  - Searchable by member demographics and capabilities, includes POC info for each member.

- Member Forum*

  - Post on the 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 govmates.com/ati.

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

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

AMERICA'S DATAHUB
CONSORTIUM

# Communications

RFS Questions: ADC-Contracts@ati.org

General/Membership Questions: adc@ati.org

**AMERICA'S DATAHUB**
CONSORTIUM

# Visit the ADC Website

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

www.americasdatahub.org

# Questions?