

## Request for Solutions:

**-Artificial Intelligence for Enhancing Data Quality, Standardization, and Integration for Federal Statistics**

**-AI-Ready Data Products to Facilitate Discovery and Use**

**Dr. May Aydin**  
NCSES, Science Advisor



AMERICA'S DATAHUB  
CONSORTIUM



## 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



# Background



AMERICA'S DATAHUB  
CONSORTIUM



# America's DataHub Consortium (ADC)

**Vision:** To be an enduring national asset, where eligible people and secure data come together for collaborative research and decision-making that will benefit the American public.

21 projects awarded since 2022



Support cutting-edge data infrastructure



Build data security capabilities to further increase privacy protections and public trust



Develop new ways of acquiring and linking data to yield valuable insights into critical issues

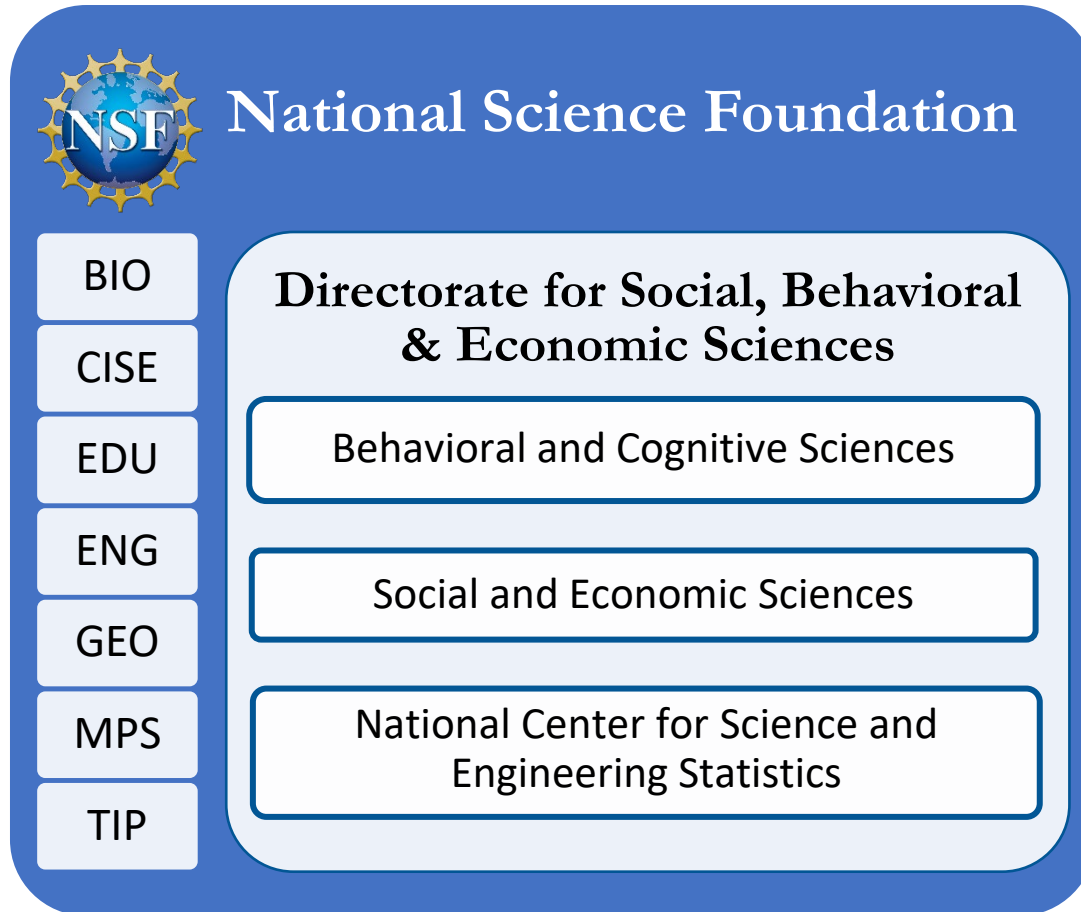


Provide novel and innovative analyses



Share lessons learned for similar activities across the Federal government

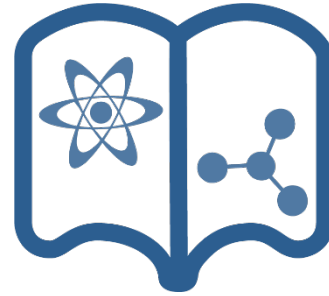
# Placed within NSF's SBE Directorate, NCSES is one of 13 principal federal statistical agencies



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



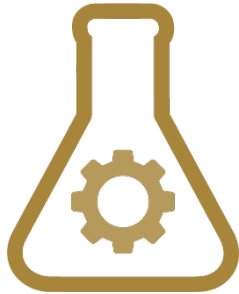
**Science & Engineering  
Workforce**



**STEM Education**



**Innovation & Global  
Competitiveness**



**Research & Development**



**Government Funding  
for S&E**



**Higher Ed R&D**



# 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

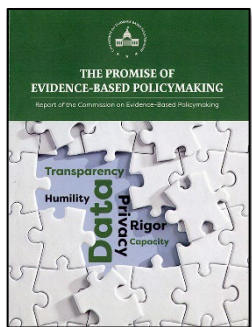
2017

2018

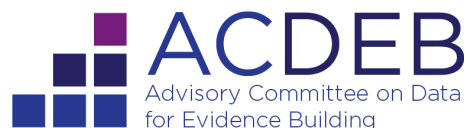
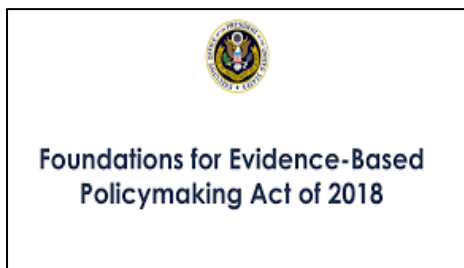
2020

2021

2022



Commission on Evidence-Based Policymaking



National Secure Data Service Demonstration Project

# A vision for the NSDS

---



The NSDS is envisioned as a government-wide set of shared services. It serves as a **front door** and a central hub for users to discover shared services and resources and to utilize the NSDS data access and linkage infrastructure.



**Shared services and resources** will include a data concierge service to direct users on their evidence-building journey, toolkits for data protection and synthetic data, communities of practice where users can connect, and platforms to promote transparency concerning the use of government data.



The **data access and linkage infrastructure** will enable users to securely access, link, and analyze powerful, high value data. The NSDS access infrastructure will complement agency enclaves and the Federal Statistical Research Data Center (FSRDC) network while addressing gaps in coverage. The NSDS will also facilitate secure data linkages in support of distinct, authorized projects.



# Artificial Intelligence for Enhancing Data Quality, Standardization, and Integration for Federal Statistics

Novin Ghaffari  
Statistician

U.S. Department of Transportation, Bureau of Transportation Statistics



AMERICA'S DATAHUB  
CONSORTIUM



# Background

Datasets face issues with consistency, integrity, and usability. Preprocessing data for quality control is often a prerequisite for statistical analysis.

Institutions are increasingly using newer, nontraditional methods of data collection. Examples include cellphone data, satellite and global positioning system (GPS) data, radio frequency identification (RFID) data, administrative records, and transaction records.

Nontraditional data have not been designed with statistical end use in mind but rather as a byproduct of other activities. Hence nontraditional data face more challenges in processing, standardization, and privacy protections.

Recent advances in AI and machine learning offer tools for handling data and deriving insights, especially from unstructured data sources. Effective AI tools that enhance data quality, standardization, and integration would expand the capacity for statistical activities.



# Key Objective

The objective of this project is to explore the use of artificial intelligence (AI) to assist in the processing, formatting, standardization, and integration of data to support activities that are integral to statistical agencies' production of high-quality statistics and data products. These efforts will be integrated to develop tools and services to support a future NSDS that supports several federal agency priorities, including capacity building, data access and sharing, and privacy protections.



# Information Gaps

Information gaps to be identified include:

- Summary of common data quality issues with traditional and nontraditional data faced by federal statistical agencies
- Description of types of traditional and nontraditional publicly available data as well as standardization procedures for the different types
- Frequently encountered data linkage and integration challenges by types of data
- Lessons and best practices that individual federal statistical agencies have learned and how might these be developed into a set of guidelines for all agencies
- AI technologies that are currently available and can address these data quality needs
- The hardware/infrastructure needs for developing the service



# Key Evidence Building Considerations

The following questions may guide the development of a service addressing data quality, standardization, and integrity:

- What data standardization issues exist and what can be addressed with AI in a toolkit?
- How can AI tools addressing data quality challenges be packaged into a service that is user-friendly, accessible, and amenable to regular updates?
- How can efforts at building a data standardization and integration service be coordinated with stakeholders?



# Objectives (1 of 4)

The objectives of this project are to:

1. Develop a framework for data standardization and integration services, including but not limited to the following topics. Additionally, outreach should be conducted with statistical agencies that have already considered these issues.
  - Identify common traditional and nontraditional data types (e.g., survey data, administrative data, cellphone data, etc.)
  - Identify 3-5 data types (to be determined) for standardization and integration needs (e.g., string formatting for text data, linking survey and nontraditional data, etc.)
  - Identify ethical and privacy concerns and risks, particularly with respect to utilizing AI to develop tools and services for data standardization and integration.
  - Identify AI applications that address data processing, standardization, integration, and privacy concerns
  - Identify computational and software tools for service deployment (e.g., python, R, etc.)



# Objectives (2 of 4)

2. Leverage or design tools and services (open source preferred) to facilitate data processing, standardization, integration, and privacy protection that can ingest traditional and nontraditional publicly available data types. Elements of the toolkit may include, but are not limited to:
  - Scraping and conversion of data from noneditable to editable formats (e.g., PDF or images to spreadsheet or tabular data)
  - Web data scraping tools to facilitate and automate data collection from websites and APIs
  - String formatting and natural language processing tools (e.g., topic clustering, sentiment analysis, etc.)
  - Data event detection such as outlier, anomaly, or change point detection
  - Traditional and nontraditional dataset merger assistance





# Objectives (3 and 4)

3. Develop documentation, use guidelines, and best practices:
  - Provide a checklist of best practices when processing and standardizing data
  - Provide users a summary of tools available and scope of applications
  - Document performance measures of AI tools (e.g., speed, computational resource usage, data size constraints, etc.) for comparison with standard approaches
  - Provide users with guidance and direction on addressing privacy concerns
  - Inform users about the applicability of data, the data context, and providing caveats about limitations with datasets, based on the data quality domains and dimensions outlined in the [FCSM Framework for Data Quality](#)
  
4. Package tools and services into an accessible service and user-friendly interface that may be included in a future NSDS but also accessed across different agencies and data platforms.





# Deliverables

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

- **A framework plan** that is informed by current activities and best practices across statistical agencies for the development of a data standardization and integration service and toolkit that leverages new or existing activities, tools, and processes that align with the FCSM Data Quality Framework.
- **A set of user-friendly tools** (i.e. toolkit) packaged into a service addressing or enhancing data processing, standardization, integration, and privacy protection.
- **A user interface** that is intuitive and readily integrated into a future NSDS and accessible to different agencies.
- **Documentation** that includes but is not limited to a set of guidelines, best practices, and caveats about using the toolkits and service.
- **Evaluation deliverables**
  - Monthly progress reports
  - Regular meetings with federal staff
  - Quarterly lessons learned
  - Final report detailing project goals and final outcomes including but not limited to tradeoffs in data quality and risk mitigation strategies



# Questions?



AMERICA'S DATAHUB  
CONSORTIUM



# AI-Ready Data Products to Facilitate Discovery and Use

**Paul Iwugo**

Chief, Office of Application Development  
Bureau of Economic Analysis



AMERICA'S DATAHUB  
CONSORTIUM



# Background

**What is AI-ready data?** Title II of the Evidence Act (or the “OPEN Government Data Act”) “requires public government data assets to be published as machine-readable data.” That is, “data in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost.” To be leveraged by generative AI technologies government data assets should not only be machine readable but machine understandable.

The Department of Commerce recently stood up the AI and Open Government Data Assets Working Group to “[develop] guidelines for publishing Commerce data that can be consumed by emerging AI technologies.” The Department describes machine-understandable, “AI-ready” data like this: “data that [are] enriched with contextual metadata and organized in interpretable standard formats.” By making its data AI ready, “AI models can then better interpret Commerce data, link them to similar data, and return accurate results from authoritative sources.” This posture of AI readiness applies to data assets held by other federal agencies and by partners throughout the data and evidence ecosystem.



# Key Objective

The objective of this project is to explore how a future National Secure Data Service (NSDS) could provide shared information and tools for making statistical data products more readily ingestible by AI technologies. These resources could support federal agencies and their partners throughout the data and evidence ecosystem as they balance the risks and rewards of leveraging generative AI to expand the reach of their statistics.



# Project Overview

The project involves two main parts:

## 1. Assessment of the accuracy and timeliness of federal statistics in generative AI tools.

- a) **Landscape analysis.** Perform a landscape analysis of the machine “understandability” of federal statistical agencies’ public data products. This should reflect a cross-section of statistical agencies (with a minimum of six agencies including those involved in the case study and developing the AI readiness tool described below), different information types (e.g., website, XLS, PDF, graphics, interactive tabling tool, API), and several of the most common generative AI engines.
- b) **Case study—Bureau of Economic Analysis (BEA).** The case study should include (1) a prioritized list of BEA data products for AI readiness transformation (developed in coordination with BEA and informed by the AI readiness assessment described above) and (2) for those data products selected as “high priority,” specific recommendations or alterations that would maintain the human-readable quality of the products and enhance their ability to be consumed by machines.



# Project Overview

## 2. AI readiness tool with prototyping and replicability testing.

- a) **Tool and prototype.** Develop a technical solution for transforming federal statistical products into machine understandable, AI-ready data. This solution should implement industry best practices for data management and present agencies' data and metadata to optimize their consumption by generative AI. The utility should take a generalized approach that could be applied to the data products of any federal statistical agency. This may be a stand-alone solution for optimizing machine “understandability” or may yield information that could be consumed by both humans and machines.
- b) **Testing and replicability.** Apply the prototype tool to data products from another statistical agency (or agencies), run tests for machine understandability (including semantics), document differences from the BEA prototype, update the tool (as appropriate), and recommend next steps for wider use.



# Information Gaps

Key outcomes of this project include the following:

- Support federal agencies in building capacity to (1) evaluate generative AI technologies' access to and use of their statistics, (2) identify high-value assets to make AI ready, and (3) provide solutions for doing so.
- Develop a tool that can be used by federal agencies to transform data assets into machine-understandable, AI ready data products.
- Lay groundwork for (1) expanding decisionmakers' access to federal data through widely available generative AI tools and (2) connecting to an AI interface tailored to respond to user prompts with authoritative statistical information, potentially as part of a future NSDS.





# Key Evidence Building Considerations

Key questions include the following:

- How is generative AI currently using trusted federal datasets to respond to user prompts?
- How can federal agencies improve their data dissemination methods and ensure the integrity, accuracy, and timeliness of their data when used in AI applications?
- How can agencies identify which data products would be most impactful if made more ingestible by AI algorithms?



# Deliverables (1 of 2)

As AI technologies rapidly evolve, it is essential that this project keeps pace with advances in the field and aligns with best practices and standards as they develop. In addition, project offerors should demonstrate AI technical expertise as well as in-depth knowledge of federal statistical products.

At a minimum, offerors will provide the deliverables described below if selected for an award. For more information, see “Project Overview.” Additional deliverables may be required.

- **Biweekly or monthly status meetings** with the project team.
- **Monthly updates** on progress toward project objectives, including quarterly highlights of lessons learned.
- **Interim reports** describing (1) the landscape analysis and case study and (2) the BEA prototype and replicability testing as well as (3) a final report that can be used as a roadmap for other federal agencies to increase AI readiness, highlighting successes, challenges, and lessons learned during this project.



# Deliverables (2 of 2)

- **An AI readiness assessment** that could be a shared resource for any agency looking to test the machine understandability of its public data products.
- **A generalizable and sustainable technical solution** that transforms statistical data products into AI-ready data and step-by-step documentation that describes how to apply the solution, using examples from the BEA and replicability tests.
- **Recommendations** on hosting and sharing reports, tools, and documentation developed in this project to encourage re-use and refinement as part of a shared service within a future NSDS.
- **Communications plan** to promote the AI readiness solution and resources with federal agencies and their data partners.



# Questions?



AMERICA'S DATAHUB  
CONSORTIUM



NCSES

# Request for Solutions (RFS) Requirements

**Alexis Hurr**

Contracts Manager

Advanced Technology International (ATI), ADC Consortium Management Firm (CMF)

[americasdatahub.org/opportunities](https://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 own risk.*



AMERICA'S DATAHUB  
CONSORTIUM



# RFS Summary

## Project Topic

- Proposals must address the specific topic area.
- See Attachment 1 of the RFS for full topic description

## Project Awards

- **Artificial Intelligence for Enhancing Data Quality, Standardization, and Integration for Federal Statistics (AI-DQSI-24):** It is anticipated there will be one award estimated at \$750,000.
- **AI-Ready Data Products to Facilitate Discovery and Use (AI-RD-24):** It is anticipated there will be one award estimated at \$1,000,000.

## Period of Performance

- **Artificial Intelligence for Enhancing Data Quality, Standardization, and Integration for Federal Statistics (AI-DQSI-24):** 18 months
- **AI-Ready Data Products to Facilitate Discovery and Use (AI-RD-24):** 12 months



## 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 form: [https://atisc.formstack.com/forms/adc ai dqsi rfs](https://atisc.formstack.com/forms/adc_ai_dqsi_rfs)
  - Submission form: [https://atisc.formstack.com/forms/adc ai rd rfs](https://atisc.formstack.com/forms/adc_ai_rd_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 POCs
- Unique Entity ID (UEI) of the Offeror (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:** 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.



# 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.

***Do not include company-sensitive or proprietary data included in the Statement of Work***



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

- **Capabilities and Experience**
  - List all project partners and indicate if they are a non-traditional entity
  - Identify all key personnel and describe their roles; organize the team by organization name
  - Relate the capabilities and experience of key personnel and organizations to the project
  - Identify any supervisory relationships and the main POC check-ins during the project
  - Provide resumes (2-page max) for all key personnel in an appendix (excluded from page limit)



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

## • Capabilities and Experience *(continued)*

- Designate any graduate students or postdoctoral fellows to be funded by the project
  - If named, provide a biographical sketch (½ page max) of their background and research interests within resume appendix
- Describe unique capabilities that may reduce risk, duration, and/or improve financial performance
- Address any potential conflicts of interest and any proposed mitigation, and complete Exhibit 1 – Organizational Conflicts of Interest Certificate



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

## • Intellectual Property Rights

- Identify limitations on any intellectual property (patents, inventions, trade secrets, copyrights, or trademarks) that will impact the performance or the Government's use of any deliverable under the project
- Describe the intellectual property in sufficient detail to describe:
  - Limitations (data assertions, potential patent licenses required by the Government, etc.)
  - Why or how the Government can accomplish the objectives with the proposed limitations



## Volume 2: Cost Proposal Content

- **Agreement Type:** Identify the preferred agreement type (e.g., firm fixed price, cost-plus-fixed-fee, etc.) and provide rationale.
  - Agreement type will be subject to concurrence of selected offeror and Government.
- **Cost Estimate:** Account for entire cost of project, broken down for each phase of the proposed work. Contractor format for the cost estimate is acceptable.



## Volume 2: Cost Proposal Content *(Continued)*

- **Labor – Offeror only:** Describe each labor category or person with labor rate and hours.
- **Travel – Offeror only:** List number of trips and number of days, travelers, and costs per trip.
- **Team Members/Subcontractors/Consultants:** List all team member/subcontractor/consultant and associated totals.
- **Material/Equipment – Offeror only:** 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:** List all items and provide justification and basis of cost for each (i.e., catalog pricing, vendor quote, previous purchase, etc.).





## Vol 2: Cost Proposal Content *(Continued)*

- **Indirect Costs – Offeror Only:** Breakout of indirect costs; indicate if indirect rates are Government approved.
  - If approved: cite approval date and federal agency.
  - If not approved: explain how the proposed indirect rates are appropriate for pricing.
- **Profit/Fee:** Indicate any proposed profit/fee.



# 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"? \*

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.



# 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?

Yes

No



# 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](mailto: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

*The criteria are listed in order of relative importance.*

- **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
  - (iii) demonstrates innovation
  - (iv) informs strategic priorities of the NSDS
- Teaming: The degree to which the proposed project includes a diverse team of qualified performers to include use of non-traditional entities.



## Full Proposal Evaluation Criteria *(Continued)*

- **Cost**

- The CMF will perform an analysis and will provide the results to the Government; 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.



# Timeline

	Dates	
	<u>AI-DQSI-24</u>	<u>AI-RD-24</u>
Request for Solutions Release	June 26	June 27
Webinar	July 1	July 1
Teaming Speed Networking Event <i>(next slide)</i>	July 2	July 2
Full Proposal Deadline	July 24, 3 PM ET	July 25, 3 PM ET
Offeror Notifications	August - September	August - September
Award Projects	September	September

*Any deadline updates will be communicated via email.*





# Teaming Resources

- **Teaming Speed Networking Event on July 2 1 PM ET**
- The event aims to help organizations find potential partners for this RFS. Each presenting organization will have a maximum of three minutes to highlight its capabilities and specify if they are looking to serve as a prime contractor, subcontractor, or either.
- Visit the “Events” page to learn more and to register for the event!

<https://www.americasdatahub.org/events/>



# Teaming Resources (continued)

- ADC Member Profile Database\*
  - Searchable by member demographics and capabilities, includes POC info for each member.
- Need a teaming partner outside of ADC or other resources?
  - Email [ati@govmates.com](mailto:ati@govmates.com) with who you are looking for.
  - More information is available at [govmates.com/ati](https://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!*



# Stay Engaged

✉ Solicitation and Contract Related Questions: [ADC-Contracts@ati.org](mailto:ADC-Contracts@ati.org)

✉ General/Membership Questions: [adc@ati.org](mailto:adc@ati.org)

✉ Join the ADC Mailing List: <https://www.americasdatahub.org/adc-mailing-request-form/>

## Visit the ADC Website

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

[www.americasdatahub.org](http://www.americasdatahub.org)



AMERICA'S DATAHUB  
CONSORTIUM

# Questions?



AMERICA'S DATAHUB  
CONSORTIUM

