

Data Access Alternatives: Artificial Intelligence Supported Interfaces (DAA) FAQ

As of May 14, 2024

	Question	Answer
1	Will slides be shared with attendees?	The presentation slides will be posted to the ADC website on the Events page.
2	What development environment will this be built in? Will it be government hosted?	It will not be government hosted at this point. If it would need to transition to that, then we would need a plan for that but the environment that it will be built in will be up to the proposer. We don't have the confidentiality or the controlled unclassified information requirements around some of our other projects. It's all public data so it is up to the proposer.
3	What if we have a patent-backed solution that accomplishes your goals? Are you open to funding/enhancing existing commercial solutions?	While it is the government's preference for an open source solution, proposals can include non-open source solutions with clearly specified benefits for the government to consider.
4	Are you seeking a Federated Join capabilities when querying the data amongst disparate datasets or is it data analysis and statistics for individual datasets.	It is up to the proposer to propose. The key objective is to increase access via a machine backed interface.
5	What is the format of the data within the data sources? (HDFS, csv, RDBMS, flat files)	These are a variety of data all available on websites. Some of it's downloadable via Excel files, CSVs, HTML, or in Pdf, so it really would be the whole website. But also, if the proposer wants to ring-fence it into a particular data type, it would be up to them, as long as they're meeting the key objective of increasing data, then we would take that into account.
6	Should proposers suggest publicly available PUF/PUMS data to include in the prototype project, or would NCSSES be specifying which data sets are included?	We will be specifying which statistical agencies are included, but the idea is that all of the data made available by the agency would be quarriable by this tool. The RFS is written so that the Cisco agency would have less than a hundred individuals. We would not, by that definition in the RFS, be using census data. Additionally, all the other websites would give a good idea of the different kinds of data available, but we also would leave it up to the proposers.
7	Do you have a target funding level for this RFS?	No, we do not. Please propose what will meet the key objective.
8	What is the floor/ceiling on the award amount?	No ceiling is defined. Please propose what will meet the key objective.

	Question	Answer
9	What is the expected output format for the data analysis?	This will need to be proposed by the proposer. Note that it would depend on what the tool being proposed is.
10	Should the solution be a plug-in to an existing .GOV website? Or is the plan to migrate the solution as a web application hosted as a .GOV	That will be left up to the proposers and what the proposed solution can do. It will eventually have to live on a .GOV site.
11	The pilot should identify how portable and scalable the existing tool is. How does the government define portable/scalable? Is this portability across other .GOV websites and scalable across many users?	The definition will be part of the project and will be one of the things that is investigated. Yes, portability is across other websites and users. If there are other interpretations to that, then it will be up to the proposers to define.
12	Do you want the ability to analyze text data or just numerical data?	The tool should answer questions related to the existing Federal statistical data, so it is up to the proposers to ring-fence.
13	What specific policies governing agency use of AI apply here? OMB guidance? Any others?	It is public information; however, the AI guidance and any contract related clause would apply. OMB has come out with some guidance, which is cited in the RFS.
14	Would the analysis need to be conducted Real time or can this be done priori - i.e.: If the same analysis was asked before, can the results be cached and reserved from Cache?	This is up to the proposer.
15	Would you consider proprietary licensing for Large Language Model (LLM) but a tool accessing it built using Open-source solution OR it must be entirely Open-source solution including LLM?	While it is the government's preference for an open source solution, proposals can include non-open source solutions with clearly specified benefits for the government to consider.
16	Is the goal explicitly about providing a singular statistic to the user or about directing interested parties towards the relevant data for more complex analyses?	The key objective is to increase the use of data; so, this will be up to the proposer.
17	What would be the maximum level of datasets that would need to be connected to - real time? (alternatively, would the end user choose a specific dataset prior to asking the question or analysis)	We are requesting solutions based on the idea of what we want our future state to be. Technical items, including tradeoffs, should be proposed in your proposal.
18	If we have any follow on questions, then what's the deadline for submitting those?	The RFS is due May 10th. The FAQ document is updated as questions are answered and approved by the AO for posting. The sooner a question is sent in, the better.
19	Is there any expectation to store and Archive resulting data based on queries from end users? (the outcome of end user queries)	This is up to the proposer. The RFS does request that "this project will identify the risk associated with the tool".
20	Is there a IGCE in terms of LOE or approximate dollar amount that you can share?	No

	Question	Answer
21	Can you share titles or personas of stakeholders the vendor will be interacting with during the project?	This is not a stakeholder engagement project; so, it is up to the proposer and the subsequent project to define those.
22	Would dynamic real time visualizations be possible or would a pull for existing graphics be more feasible?	This is up to the proposer.
23	Will the NSF Secure Compute Environment Testbed for a National Secure Data Service (SCET) opportunity be awarded before this opportunity is awarded/funded?	This is not known at this time.
24	Are foreign components allowed?	This is up to the proposer to discuss the risk associated with the tool. Note that the public websites are available publicly and everything has risks.
25	Is the expectation that queries return existing graphics or create dynamic visualizations on demand?	This is up to the proposer.
26	Are proposers' future phases desired to include a FedRAMP-secure environment as the initial effort is envisioned as publicly accessible?	Responders should propose the benefits of their proposed solution.
27	Would you be amenable to entire solution to being T&M vice FFP?	Historically, ADC has done firm-fixed-price or cost-plus-fixed-fee contracts. Proposers can propose their preferred contract type, but FFP & CPFF are preferred.
28	How many users are expected to visit and interact?	This request focuses on a proof of concept, not a public tool. The solution should be scalable, the parameters of which should be detailed in the proposed solution.
29	Can you provide guidance on how important having multiple teaming partners are for technical evaluation? Is there any downside if you bid as a single firm?	Teaming is strongly encouraged but firms may submit proposals without partners.
30	Can you speak about Non Traditional teaming partners in evaluation?	Teaming, including with non-traditional partners, is strongly encouraged. This is defined in the base agreement.
31	The government indicates that "the tool should be developed using or resulting in open-source software or code..." Does this mean that the proposer must propose an open source foundational LLM, or can the proposer provide a licensed, commercial LLM that all other code, built on top of the foundational LLM is open source, but the core infrastructure/LLM is considered COTS"?	While it is the government's preference for an open source solution, proposals can include non-open source solutions with clearly specified benefits for the government to consider.
32	Is the data hosted on multiple websites? Or by multiple agencies?	Yes.

	Question	Answer
33	Are you considering multiple awards if funding is available?	Yes
34	How is the three-year period defined when considering whether an entity is non-traditional? For example, if an entity had an NSF contract that ended <i>within</i> 3 years, but it began <i>more than</i> 3 years ago would that entity be considered non-traditional or not?	The definition states “that is not currently performing and has not performed. So if the entity worked on an NSF contract anytime within that preceding 3-year period prior to the RFS release, then they would be a traditional.
35	Will performers be provided with access to any SMEs or similar points of contact at NCSES or other statistical agencies involved in the pilot project? Will this access include the ability to perform interviews/surveys?	Yes. Please propose information gathering needs in your solution.
36	How many statistical agencies will be included in the pilot project? How will they be selected (i.e. by the respondent or by NSF?)	NSF/NCSES, in conjunction with other ICSP members, will select 3 agencies to be in scope for the pilot.
37	Should the requested transition plan be focused on deploying the solution on each agency's (.gov) website? Or on a new standalone (.gov) website?	Both.
38	Can respondents suggest inclusion/exclusion criteria for the types of public data that should be included in the pilot?	Yes.
39	Will respondents be provided with any back end/API-based access to public agency data? Or will respondents be required to manually download data used in the pilot?	Respondent will need to acquire the data from the public websites.
40	Is NSF prioritizing specific user communities for the pilot (e.g. federal employees, citizens, researchers, state/local governments)?	No
41	Are there any restrictions on the staff allowed to perform work on this contract (i.e. US citizenship, permanent residency, H1-b visa?)	Yes. All work must be performed by staff residing within the U.S.
42	Does the data concierge service mentioned in the RFS refer to a specific function/business unit within the NSDS that includes human headcount? Or is the data concierge expected to be an automated/software process?	This has not been defined yet.
43	Is this solution expected to be interoperable with other federal open data tools, like data.gov?	Not at this stage.
44	What, if any, non-traditional search engines (i.e. excluding Google, Bing) are currently used to answer questions about or search for public data?	This is unknown.
45	Can the proposed solution integrate any proprietary or closed source Large Language Models (e.g. Google Gemini, OpenAI's GPT4)? Or should any Large Language Models used in the solution be open-source?	While it is the government’s preference for an open source solution, proposals can include non-open source solutions with clearly specified benefits for the government to consider.

	Question	Answer
46	Does the UI required analysis and reports or is a UI that only "Pulls" Information?	Respondent should propose the best solution to meet the key objectives of the RFS.
47	ATT-2_Full-Proposal Format Page 1 Is there a preference or requirement for font type and size?	Full proposals shall be no less than 10- point font. No font type is required.
48	ATT-2_Full-Proposal Format Page 3 If the offeror proposes firm fixed price, in that case it is our understanding that individual elements listed in the Program Funds table (i.e. overhead, profit etc.) required to be detailed separately in the cost proposal?	The cost proposal format and preparation format instructions apply to all proposed agreement types.
49	ATT-2_Full-Proposal Format Pages 3 & 4 The Cost Share portion of the table includes "(not required)" yet the instructions above the table states "The cost estimate shall account for the entire cost of the project, inclusive of any proposed cost share." - please confirm whether non-traditional offerors need to address the cost share elements in their cost proposals.	Respondents not planning to utilize or propose cost share do not have to address the cost share elements in the cost proposal. Please reference the definition of a nontraditional as indicated 3.1.1.2 of the RFS.
50	ATT-1 Pages 2 & 3 How many simultaneous users should the solution support?	This request focuses on a proof of concept, not a public tool. The solution should be scalable, the parameters of which should be detailed in the proposed solution.
51	ATT-1 Pages 2 & 3 Do users need to be able to authenticate to the system before use?	Respondent should propose the best solution to meet the key objectives of the RFS.
52	ATT-1 Pages 2 & 3 How much data (xx gigabytes) does the solution need to store?	Respondent should propose the best solution to meet the key objectives of the RFS.
53	ATT-1 Pages 2 & 3 What are the personas/types of users using the application?	The proposer should assume a wide variety of users of and needs for federal statistical data.
54	ATT-1 Pages 2 & 3 What is the estimated total user count and concurrent user count?	This request focuses on a proof of concept, not a public tool. The solution should be scalable, the parameters of which should be detailed in the proposed solution.
55	ATT-1 Pages 2 & 3 What constitutes "high accuracy" in the context of returned statistics, and how should this be quantified?	Respondent should propose the best solution to meet the key objectives of the RFS.
56	ATT-1 Pages 2 & 3 Should the AI tool generate graphics dynamically based on data, or retrieve pre-existing graphics from documentation? What formats or tools are preferred for these graphics?	Respondent should propose the best solution to meet the key objectives of the RFS.
57	ATT-1 Pages 2 & 3	Respondent should propose the best solution to meet the key objectives of the RFS

	Question	Answer
	Will this solution generate analyses and draw inferences between documents, or does it simply display specific excerpts of text and data sourced directly from the documents and internet?	
58	ATT-1 Pages 2 & 3 Will this solution be expected to source real-time data on agency websites?	Respondent should propose the best solution to meet the key objectives of the RFS.
59	ATT-1 Pages 2 & 3 What file extensions will be used for data files?	These are a variety of data all available on websites. Some of it's downloadable via Excel files, CSVs, HTML, or in Pdf, so it really would be the whole website. But also, if the proposer wants to ring-fence it into a particular data type, it would be up to them, as long as they're meeting the key objective of increasing data, then we would take that into account.
60	ATT-1 Pages 2 & 3 Which specific federal statistical agencies' data should be considered for inclusion beyond NCSES? Are there predetermined agencies or will the proposer need to select them based on relevance and availability?	This is up to the proposer.
61	ATT-1 Pages 2 & 3 Are there particular formats of data (e.g., demographic statistics, economic indicators, health statistics) that the project should focus on using?	Respondent should propose the best solution to meet the key objectives of the RFS.
62	ATT-1 Pages 2 & 3 Are there existing accuracy benchmarks or standards within NCSES or other federal statistical agencies that the proposal should align with?	No.
63	ATT-1 Pages 2 & 3 What specific federal or agency-specific policies regarding the use of AI need to be considered?	The AI guidance and any contract related clause would apply. OMB has come out with some guidance, which is cited in the RFS.
64	ATT-1 Pages 2 & 3 About how many users do you expect to use the system per day and about how many questions do you expect each user to ask and how long do you expect each question and response to be in terms of words? We typically see a question of about	This request focuses on a proof of concept, not a public tool. The solution should be scalable, the parameters of which should be detailed in the proposed solution.
65	ATT-1 Pages 2 & 3 About how many total users do you expect to use the system and about how many questions do you expect each user to ask.	This request focuses on a proof of concept, not a public tool. The solution should be scalable, the parameters of which should be detailed in the proposed solution.
66	ATT-1 Pages 2 & 3 How long do you expect each question and response to be in terms of words? We typically see a question length of about 20 words and a typical half-page response of about 250 words.	Respondent should propose the best solution to meet the key objectives of the RFS.
67	ATT-1 Pages 2 & 3 Will the government be amenable to using COTS products that are NOT open-source be proposed as part of the solution? Typically,	While it is the government's preference for an open source solution, proposals can include non-open source solutions with clearly specified benefits for the government to consider.

	Question	Answer
	COTS products would NOT share source code on GitHub repository.	
68	ATT-1 Pages 2 & 3 Would government be amenable to offeror developing an application/tool as a solution using open-source code but utilizing proprietary Large Language Model (LLM) to generate the response. The offeror would be publishing the application/tool developed as part of this RFS but NOT the LLM source-code as that is 3rd party proprietary code.	While it is the government's preference for an open source solution, proposals can include non-open source solutions with clearly specified benefits for the government to consider.
69	ATT-1 Pages 2 & 3 What FedRAMP classification (Low, Moderate, High) do you anticipate once the website is migrated to a .gov domain?	Responders should propose the benefits of their proposed solution.
70	ATT-1 Pages 2 & 3 What specific types of user support or help desk services will be necessary to effectively manage this solution?	This is unknown at this time.
71	ATT-1 Pages 2 & 3 What are the requirements, if any for training materials and sessions?	Respondent should propose the best solution to meet the key objectives of the RFS.
72	What are the expected volumes of user queries, and does the system need to scale to meet varying demands?	This request focuses on a proof of concept, not a public tool. The solution should be scalable, the parameters of which should be detailed in the proposed solution.
73	Who are the primary and secondary users of the chatbot, this might shed some light on their specific needs?	The proposer should assume a wide variety of users of and needs for federal statistical data.
74	Is there a preference in the visualization platform or would the source of response to the query dictate that visual presented?	Respondent should propose the best solution to meet the key objectives of the RFS.
75	Are there any federal data protection and privacy laws that the chatbot must comply with?	Yes. The AI guidance and any contract related clause would apply. OMB has come out with some guidance, which is cited in the RFS.
76	Is there a preference for specific statistical agencies and/or bureaus that would be used in the early phases of the project?	The RFS provides guidance.
77	Is there a phased rollout planned for the chatbot? Or is it up to the contractor?	This request focuses on a proof of concept, not a public tool. The solution should be scalable, the parameters of which should be detailed in the proposed solution.
78	To assist vendors with their cost estimation for the price proposal, how many federal statistical agency website/platforms data source assumptions should be factored for the base year (2-3)?	Respondent should propose the best solution that focuses on the success of a proof of concept.

	Question	Answer
79	Is the development of the chatbot tool a stand alone requirement for a single public-facing statistical website or should multiple iterations be factored for this initiative?	Respondent should propose the best solution to meet the key objectives of the RFS.
80	ATT-1 Pages 2 & 3 Can government please clarify if offeror is expected to generate data/tables in chatbot response based on content shared or simply extract the data/tables from the statistical document shared. We are trying to ascertain if AI needs to perform statistical computations to generate response or simply present the pre-computed outcomes provided in the documents shared?	The AI should be used to retrieve and extract data/information from statistical agencies, not generate or recalculate new data/information.
81	Will the tool be generating statistics or retrieving and delivering the existing statistics (from other Federal sites) in the tool?	The AI should be used to retrieve and extract data/information from statistical agencies, not generate or recalculate new data/information.
82	Does the government want to leverage novel quantitative transformations to deliver the statistic(s) in the tool?	Respondent should propose the best solution to meet the key objectives of the RFS.
83	If we are producing novel statistics - what error threshold is acceptable converting text statistic to a graphic?	The AI should be used to retrieve and extract data/information from statistical agencies, not generate or recalculate new data/information.
84	Does the tool need to generate Visualizations (i.e. Chart and graphs)) based on the retrieval of the statistic? If so, using software do that acceptable?	Respondent should propose the best solution to meet the key objectives of the RFS.
85	Would the graphics/image being rendered in a PDF be acceptable?	Respondent should propose the best solution to meet the key objectives of the RFS. Responders should propose the benefits of their proposed solution.
86	Will the government be providing inferencing resources during the deployment of the tool?	This request focuses on a proof of concept, not a public tool. The solution should be scalable, the parameters of which should be detailed in the proposed solution.
87	Do you have an ATO process in place? Or do you expect us to get the ATO? (e.g. Limitations to integrate the tool into NSF environment)	This request focuses on a proof of concept, not a public tool. The solution should be scalable, the parameters of which should be detailed in the proposed solution.
88	Do you expect the solution to be in the Cloud? Does the government expect to architect the solution to a standard specification? (e.g. applying a STIG)?	Respondent should propose the best solution to meet the key objectives of the RFS. Responders should propose the benefits of their proposed solution.
89	Can the government provide more information about the consistency of federal policies governing agency use of AI? (refer to page 2 of Project Topic)	Responders should reference existing federal AI policies outlined in the RFS and elsewhere.
90	Can the government share a list of websites (or examples) that the tool is going to integrate with?	The RFS provides guidance.
91	Is the government able to support a third party search provider? Does the government have a unified search API that can be integrated with the tool?	Respondent should propose the best solution to meet the key objectives of the RFS.
92	Are there any metrics which NSF would like the proof of concept to be measured against? i.e., What constitutes a successful prototype in terms of generation quality?	Respondent should propose the best solution to meet the key objectives of the RFS. Responders should propose the metrics and their benefits of their proposed solution.

	Question	Answer
93	What are some evaluation metrics or benchmarks NSF would prefer to measure effectiveness and accuracy of chatbot responses?	Respondent should propose the best solution to meet the key objectives of the RFS. Responders should propose the metrics and their benefits of their proposed solution.
94	Would NSF consider an extension to 5/24?	This RFS was previously extended to 5/17. We do not intend to extend it again to ensure current milestones and timelines are met for this fiscal year.