Bidders Questions and Answers Form

RFP Reference:	IPA/RFP/DSE/011/2024
Questions Due:	18/09/2024
IPA Response Due:	20/09/2024

Question 1:	What are the most critical outcomes you expect from this project? How will you evaluate the success of the selected vendor?
Answer:	Success will be evaluated based the collaborative experience of working with the selected vendor, and the code (and supporting documentation) delivered for the infrastructure management that clearly defines how IPA can deploy, update, and maintain environments for research data management that meet IPA's data science, research, security, and governance needs within an affordable framework.
Question 2:	Can you provide more detailed information on the current state of your data infrastructure, particularly any challenges or limitations you are currently facing?
Answer:	Data currently resides on survey data collection servers and is downloaded to individual laptops and files are synced to cloud storage (Box) where it is encrypted with IPA keys. Much of the burden is placed on individual staff members relying on fairly brittle internet connectivity and bespoke data processing scripts. This introduces several potential fault points and manual data workflows and does not provide the ability to compare data across projects.
Question 3:	Could IPA please provide additional detail regarding the number and scope of data sources to be integrated? What is the expected size and complexity of the datasets to be processed through the ELT pipelines?
Answer:	The primary data sources will be data sourced from a survey API (SurveyCTO) as well as operational data from a SQL Server. Volume of data for the development environment will be minimal (MBs) that will allow for rapid prototyping.
Question 4:	Who are the primary stakeholders for this project, and what are their roles?
Answer:	Research and data science; Technology team
Question 5:	Have you selected a Cloud Data Warehouse vendor, or is that part of the scope of work?
Answer:	Selection of the cloud data warehouse is not part of the scope of work.
Question 6:	What level of internal resources and technical expertise do you have available to support this project on an ongoing basis? Is training required as part of this proposal?
Answer:	IPA will provide project management as well as technical expertise as it relates to data sources to be included in the project and authentication/security consideration. No training is required as part of the proposal; however, we do require clear, detailed documentation

Innovations for Poverty Action
1701 Rhode Island Ave NW, 3rd Floor
Washington, DC 20036

Twitter
LinkedIn
Facebook
YouTube

www.poverty-action.org

1

	as well as handoff that will enable IPA to maintain and elaborate the codebase for the development environment.
Question 7:	Are there any particular timelines or deadlines associated with key deliverables that we should be aware of?
Answer:	We expect fully operationalized development environment and code to manage the infrastructure by mid-December 2024. In the lead up to that, we will work on an agreed cadence to develop, test, and review the infrastructure code base.
Question 8:	What security and compliance frameworks must this project adhere to, if any (e.g., GDPR, HIPAA, FedRAMP)?
Answer:	For the current project, compliance frameworks are not required. However, the infrastructure code should allow IPA to set policies determined by the location for which the data is relevant, e.g. region-specific storage> We prefer the code to be extensible for potential future compliance needs.
Question 9:	Are there any specific data handling or access control policies that need to be incorporated into the data pipelines and workflows?
Answer:	Data need to be encrypted in transit and at rest. Role based access controls should conform with IPA's authentication groups and policies (e.g. determined via Entra)
Question 10:	Does IPA currently have an Active Directory, Entra ID, Okta, or other identity service provider that will need to be integrated with the Data Warehouse solution?
Answer:	Entra ID
Question 11:	Would you be open to including an upfront assessment as part of the scope of work to identify, prioritize, and plan the highest-value tasks and ensure the selected tools/platforms meet IPA's long-term needs?
Answer:	No, we seek extensive upfront assessment. We have a preferred tech stack but seek support in setting up the data governance, access, and storage via IaC.
Question 12:	Can you provide more details about the decision-making process for tool and platform selection (e.g., AWS vs. Azure, dbt vs. SQLMesh)? Is this open to guidance from the selected vendor?
Answer:	We are open to input and guidance on AWS vs. Azure. The selection of the data transformation framework is outside of this scope of work (Secure Cloud Computing Dev Environment.
Question 13:	Is there an existing budget in place for the total cost of ownership of selected tools and platforms beyond the budget allocated to this RFP? If so, are there any additional details you can share about budget constraints on the selected tools/platforms?
Answer:	We have selected solutions specifically to reduce ongoing costs of cloud computing resources and maintenance.
Question 14:	Can you provide examples of the analyses currently being conducted and the nature of the data used to drive those analyses?
Answer:	Data are collected via digital data collection software (e.g. SurveyCTO, ODK, Kobo) or come from operational data sources. The focus of the scope of work is to set IPA up for success with ELT data workflows.

Innovations for Poverty Action
1701 Rhode Island Ave NW, 3rd Floor
Washington, DC 20036

Twitter
LinkedIn
Facebook
YouTube

www.poverty-action.org   **2**

Question 15:	What type of users will have access to the database after go-live and how many total users (named and / or concurrent) do you expect?
Answer:	We see 3 groups: - Core data management, data science, and data administration staff / developers who will develop solutions, analyze across data sets, and develop business logic / ETL pipelines / manage infrastructure - about 15-30 people over time with role-based security - Research staff who need to use the tools provided and perform data analysis and management on their specific project data set - about 200-300 people - Potential for external parties in a later phase (e.g. researchers that we partner with) - out of scope initially We prefer a named user model for security and compliance purposes but are open to other models.
Question 16:	Should the implemented solution support different levels of access and / or permission sets to access to different datasets (specifically sensitive data)?
Answer:	Yes
Question 17:	Would IPA be interested in proposals which include development support that extends beyond the December timeframe that the RFP specifies?
Answer:	Yes but costed separately as different line items so we can compare directly.
Question 18:	Does IPA expect the technologies implemented for this project to support other initiatives in the future? If so, can you share what programs might be supported or what those future needs might include?
Answer:	Yes in later phases. Operational data management, reporting, and data flows.
Question 19:	Does IPA have preferences around staffing (onshore, nearshore, blended)?
Answer:	No
Question 20:	Does IPA have any preferred contract vehicles, aside from purchasing directly from chosen vendors / suppliers, when purchasing technology and implementation services?
Answer:	No

END

Innovations for Poverty Action
1701 Rhode Island Ave NW, 3rd Floor
Washington, DC 20036

Twitter
LinkedIn
Facebook
YouTube

www.poverty-action.org

3

2nd Set of Questions - Bidders Questions and Answers Form

RFP Reference:	IPA/RFP/DSE/011/2024
Questions Due:	18/09/2024
IPA Response Due:	20/09/2024

Question 1:	Can you please elaborate on the "preferred tech stack" referenced in the previous responses? What are the components of this tech stack? If selection of the Cloud Data Warehouse is not part of the scope of work, what Cloud Data Warehouse have you selected (i.e. Snowflake)?
Answer:	For local development, we will use DuckDB. For cloud data warehouse, we are considering Snowflake, Motherduck, Databricks, and Clickhouse. The data transformation is not part of this scope of work, thus the choice of CDW is not central. What is important is that we establish a cost-minimizing data lake solution that will be straightforward to connect to a CDW well equipped to ingest and compute transformations on top of object storage.
Question 2:	What tool is expected to be used for Infrastructure as Code (i.e. Terraform, Pulumi, etc.)?
Answer:	We are open to suggestions with an eye towards IPA's long-term maintenance of the codebase. Our preference is to use Python SDKs, followed by YAML, if Python SDK is not supported.
Question 3:	Do you have an existing vendor or tool selected for movement of data from the Survey API (SurveyCTO), the SQL Server, and Box API?
Answer:	We will work with Python APIs for each data source.
Question 4:	Where is the SQL Server currently located (i.e. on-premise, Azure, AWS)?
Answer:	Azure
Question 5:	In the current data infrastructure, it sounds like data transformation scripts are being executed in team members' local environments. Is part of the current scope of work standardizing and productionalizing these scripts in the cloud environment using infrastructure as code?
Answer:	Not at this stage. Our focus with this scope is to establish the Extract-Load portion of the workflow such that access policies and governance are set up properly for security and compliance.

END

Innovations for Poverty Action
1701 Rhode Island Ave NW, 3rd Floor
Washington, DC 20036

Twitter
LinkedIn
Facebook
YouTube

www.poverty-action.org

4