ATTACHMENT I – PROJECT TOPIC

Privacy Preserving Technologies
Phase 1 - Environmental Scan

Key Objective

The objective of this project is to understand the current landscape of privacy preserving technologies (PPT) for the protection of persons, data, and systems that contribute to the use of confidential data, including both individual-level and business data, for evidence-building and policymaking. For the purposes of this project, "PPT" includes technologies, techniques, methodologies, approaches, tools and other like terms that relate to preserving privacy.

Key Evidence Building Considerations

America's DataHub Consortium brings together capabilities and infrastructure to securely fill information gaps and to take on key analytic questions and evidence building challenges. As demand for access to confidential federal data assets increases alongside novel analytical approaches, privacy protections must be in place to ensure the protection of privacy and the confidentiality of the data. PPT represent a means to reduce the disclosure risk when allowing access to confidential data, and balancing disclosure risk against the need for accurate, granular output that can be used for decision-making for public policy and programs.

This project is two-fold, with the first phase providing an environmental scan of PPT currently being developed, tested, and utilized across environments and sectors, and the second phase using this information to pilot one or more of these PPT in a real-world research setting.

Phase 1 – PPT Environmental Scan

This portion of this project will produce an overview of pilots and projects currently testing or implementing privacy preserving technologies throughout government, academia, and the private sector. The deliverable would focus on the topics below:

- What projects and pilots are currently testing or implementing (or have previously been done) privacy-preserving technologies, including but not limited to secure multi-party computing (SMPC), synthetic data, differential privacy methodologies, homomorphic encryption and validation servers?
- What lessons learned are available from using PPT, including what has worked/is working and under what contexts/purposes/various types of data users, what challenges or barriers have been discovered with PPT in using these technologies from the data provider and data user perspectives, and what potential next steps are there in implementing these technologies?
- What do we know about how to evaluate whether a certain PPT is a good fit for a specific use case? What features of the data or users might indicate one PPT approach over another?
- What are best practices for effective communication strategies and/or user training on how to conduct research or program evaluation projects that leverage these new PPT?
- What use cases exist for each of these technologies when applied to evidence-building research, policymaking, and program evaluation?

At a future date, a separate solicitation will be released for Phase 2, PPT Pilots. This portion of the project will test one or more PPT potentially using a 'traditional' approach, such as tiered access, in tandem with the PPT to test relative efficacy, accessibility, and feasibility of using the technology in a research and evaluation settings.