Attachment 1 – Project Topic

Linking State and Federal Data to Strengthen Federal Program Evaluation

Key Objective

Evaluation is a key agency function. The ability to link administrative data records across multiple sources is critical to understanding results achieved with federal funding. Many federal programs are administered at the state level, but the federal government does not have or has limited direct access to the administrative data collected by the states from these programs (examples of such programs include Medicaid, Supplemental Nutrition Assistance Program (SNAP), unemployment insurance (UI), and Temporary Assistance for Needy Families (TANF)). Although these state data are valuable for understanding the impact of federal programs, there are numerous, well-documented barriers to accessing and using them.

This Request for Solutions (RFS) seeks to develop a toolkit or framework for the federal community to facilitate federal-state data linkages for evaluation activities by: (a) conducting an analysis of current federal-state linkages in a federal program evaluation office to identify implications for other federal program evaluation offices; (b) applying innovative strategies to address data access and integration challenges by conducting a two-state use case that acquires and links state child care data with the Department of Housing and Urban Development's (HUD) Child Care and Development Fund (CCDF) block grant data; and (c) documenting solutions that can support a scalable and repeatable process for future state-federal data integration efforts.

Background

The Department of Housing and Urban Development (HUD) conducts rigorous evaluations to support the Department's mission of creating sustainable communities and expanding affordable housing. A limitation to HUD's evaluation capacity is that there is no single source that identifies the extent to which HUD-assisted households receive benefits and/or support from multiple government programs. For example, HUD wants to gain a better understanding of child care needs, preferences, and access among HUD-assisted families. Obtaining a better understanding of the extent to which HUD recipients are receiving child care vouchers may inform HUD's efforts to promote economic independence among HUD-assisted households.

Child care challenges are significant for low-income families, potentially limiting the ability of parents to search for, obtain, and maintain jobs. However, federally supported child care programs are administered at the state level and recent attempts by HUD to form partnerships with states to identify HUD assisted families receiving child care vouchers have been unsuccessful. This issue is not a one-time problem; similar challenges arise as HUD routinely requests access to administrative data managed by states to assess how HUD is serving its target population. Furthermore, challenges with federal-state data linkages are not a HUD-specific challenge. Other federal evaluation agencies have attempted to address these challenges by partnering with state agencies; however, these efforts are hindered by existing regulatory barriers, limited access, lengthy processes, and other challenges.

With this project, HUD is proposing to serve as a case study to inform the work of other federal evaluation offices also facing difficulties in partnering with state agencies to access important administrative microdata, strengthen evaluation activities, and create evidence for decision-making. By implementing this project through the NSDS, we can learn how to streamline this process and provide evidence and data gathered through linked datasets to a wider audience, both technical and non-technical.

The Current Request for Solutions

The toolkit or framework will be informed by two elements: (1) linkage analysis and (2) use case. First, the contractor will conduct a **linkage analysis** of how the process and outcomes of current HUD-state linkages have served the needs of HUD evaluation, identifying and assessing the implications for evaluation offices across government agencies. Second, the contractor will employ a **use case** developed by HUD to pilot in the NSDS secure compute environment testbed. Operationalizing linkages between HUD and state administrative data for evaluation purposes will inform the development of a replicable system that could be used by other federal agencies with similar needs.

Linkage Analysis
For the analysis, the project team will work with HUD stakeholders to identify relevant contacts and data sources. The analysis will address, but not be limited to, the following topics and research questions:

- How can data linkages with state partners improve and expand the evaluation functions of HUD?
- What research questions can only be addressed by linking HUD administrative data with state-level data?
- How has HUD's current data and IT infrastructure affected its ability to execute partnerships with state agencies and securely link data?
- What lessons from HUD's previous data-linkage partnerships with state agencies are most relevant for evaluation offices across the government?

Use Case
For the use case, the project team will collaborate with state agencies to gain access to child care data that will allow HUD to answer the following research question: To what extent are HUD-assisted households receiving child care vouchers funded by the Child Care and Development Fund (CCDF) block grant?[1] This is built on past efforts by HUD to learn more about child care needs of HUD assisted households.[2]

The use case will involve two states. Subject matter experts at HUD will work with the project team to identify states that could be contacted to participate in the project. The project team will execute a data agreement and implement a data linkage with the first state (State 1) and replicate the process with the second state (State 2). As such, the use case will involve the following key activities:

---

[1] For more information on CCDBG, please see https://www.acf.hhs.gov/occ/fact-sheet

[2] The Consolidated Authorizations Act, 2021 authorized a collaboration with HUD and the Department of Health and Human Services to promote childcare services to HUD assisted families. In preparation, HUD funded a study to learn more about childcare needs of public housing residents.

- Collaborating with HUD to identify State 1 and State 2 for implementation and replication of the use case.
- Developing and completing a data agreement with State 1, with the intention of developing a template agreement that can be used for future acquisitions of state program data.
- Executing and documenting the data ingestion and data linkage process in the NSDS secure compute environment testbed for State 1, with a focus on standardization, replicability, and privacy preservation.
- Replicating and documenting the data ingestion and linkage process for State 2, making necessary adjustments to the approach, identifying issues encountered, and building flexibilities to account for the differences between states' systems with the primary goal of simplifying future data updates and application to other states.

Findings from the use case can inform the resources, tools, and infrastructure necessary to facilitate data partnerships between federal and state agencies within a future NSDS, enhancing the evaluation capacity of the federal government.

Information Gaps

This project will:

- Allow federal agencies to obtain a more complete understanding of the utilization of state-administered government programs, which could enhance programs and policies in the long run.
- Allow program evaluators to ask and answer more complex research questions and better estimate impacts as federal-state data linkages provide evaluators with a better understanding of the characteristics of low-income population.
- Develop a linkage framework to support secure integration of federal and state administrative data for program evaluation efforts.
- Identify lessons learned in the effort to analyze, acquire, and use administrative data sources at various levels (e.g. federal and state) and assess data sources fitness for use to inform program evaluation efforts using the secure compute environment testbed within the NSDS.

Key Evidence-Building Considerations

This project will inform the following questions:

- What role can the NSDS play to streamline the administrative process required for state-federal data linkages?
- Can the NSDS serve as a trusted third-party, independently facilitating state-federal data linkages?
- How would program evaluators interact with a future NSDS? How could the case study inform the future NSDS data concierge service?

Deliverables

At a minimum, the contractor will provide the following deliverables. Additional deliverables may be required.

- Monthly status reports on progress towards project objectives.
- Quarterly lessons learned based on what has been learned during the last quarter that will inform a future NSDS.
- Bi-weekly check-ins with the AOR and the federal technical team, including identifying lessons learned on a quarterly basis.
- Separate draft data collection and analysis plans for the linkage analysis and use case. Submission of final plans following review and input by the AOR and federal technical team.
- Source code, data dictionary, crosswalks, and other applicable documentation to replicate data linkage and ingestion process.
- Linked dataset ready for analysis in the NSDS secure compute environment testbed and accessible to HUD.
- A report that includes description of linkage process, description of coverage of the linked data, a summary of the quality of resulting linked data (e.g., estimates of precision and recall), and assessment of potential linkage bias, including information on how the use case performed within the NSDS Secure Compute Environment Testbed.
- Three final reports:
  - Linkage analysis, including but not limited to methods, findings, and recommendations.
  - Use case, including but not limited to data infrastructure, methods, data agreement, and recommendations.
  - Lessons learned from the project of a length and style suitable for the general public.
- A toolkit, framework, and/or other proposed tool that provides a user-friendly guide for the federal evaluation community to facilitate federal state-data linkages for evaluation activities.

Privacy Requirements

The National Center for Science and Engineering Statistics (NCSES), within the U.S. National Science Foundation, is a federal statistical agency with authority in law and expertise in practice to protect Americans' confidential data. The CHIPS and Science Act, Section 13075, positioned the NSDS demonstration project at NCSES with clear mandates about privacy protections. It specifically called for NSDS to protect "confidential data and statistical products" and to ensure that "no individual entity's data or information is revealed by the National Secure Data Service demonstration project platform to any other party in an identifiable form." Any datasets resulting from linkages with restricted data have the same protection as the restricted data. Linking restricted data with a public dataset, for example, results in a linked dataset that is itself restricted.

All NSDS projects using restricted data must be performing statistical activities for statistical purposes. These terms are defined as follows:

- *Statistical Activity* – The collection, compilation, processing, or analysis of data for the purpose of describing or making estimates concerning the economy, society, or the natural environment as a whole or relevant groups or components therein.

- *Statistical Purpose* – The description, estimation, or analysis of the characteristics of groups, without identifying individuals or organizations that make up such groups.