Federal Statistical Research Data Center (FSRDC) Executive Directors' response to the Request for Comment from the Advisory Committee on Data for Evidence Building

Please accept the following responses to the set of 10 questions included in your Federal Register Notice (https://www.federalregister.gov/documents/2020/12/15/2020-27489/request-for-comments-for-the-advisory-committee-on-data-for-evidence-building). They are submitted jointly by the executive directors of the 32 FSRDCs around the country. I have organized the material in order according to those questions.

Sincerely,
Wayne Gray, Executive Director, Boston Federal Statistical Research Data Center

For the FSRDC Executive Directors - February 9, 2021

1. What are the main challenges faced by national, state/provincial, or local governments that are trying to build a basis for evidence-based policy? Briefly describe the bottlenecks and pain-points they face in the evidence-based decision-making process.

Few agencies have their own pool of skilled researchers able to employ the sophisticated statistical techniques that would produce valid causal inferences for evidence-based decision-making. However, agencies are understandably reluctant to provide access to their data for researchers from outside the agency unless they can be sure the data will be protected. Years of experience may be needed to build the trust of agencies to allow their data to be used. The FSRDCs provide a cost-effective way to engage the services of the academic community for these evidence-building efforts, as those services are provided at no direct cost. The ability for the academic community to provide these services depends critically on their being sufficient support to maintain the FSRDC network, including the staff of both federal agencies and institutional partners who support researchers in understanding, accessing, and working with restricted federal data resources.

In addition, many projects require data from multiple agencies. Navigating the administrative hurdles across agencies and between layers of government is one of the most difficult challenges for building datasets usable for evidence-building. Agencies providing data operate under different legal restrictions and can have different interpretations of the same statutes. Projects that combine data from multiple sources are especially vulnerable because everyone needs to say "yes" before the project can proceed. The effort required to build a supportive coalition of agencies is often prohibitive for any individual project. Further complications may arise if projects require data to be obtained from non-government sources. Clear safeguards that allow data sharing for evidence building, while prohibiting sharing that would infringe on privacy or allow the use of data for enforcement purposes, would make valuable evidence-building projects possible.

2. What are examples of high-impact data uses for evidence-based policy making that successfully effected change, reduced costs, or improved the welfare of citizens?

Evidence-based policy making relies on both basic and applied research. Basic research provides a better understanding of the impacts of policies in general and how individuals and businesses respond to those policies. This is often carried out by academic researchers advancing the borders of scientific knowledge. Applied research is targeted at particular policy decisions and may be carried out by consultants and contractors working for the agencies to support their policy deliberations.

Over the years, researchers working in the FSRDCs have conducted hundreds of research projects with important implications for policy – focusing on trade, productivity, and health among many other topics. Many of these projects are described in the RDC Project Metadata at https://www.census.gov/about/adrm/fsrdc/about/ongoing-projects.html. These have generally been of the basic research type and the vast majority have involved academic researchers or academic researchers working with scientists in statistical agencies, the Federal Reserve System, and federal agencies.

FSRDC projects have included researchers and analysts from consulting firms and contractors. Both the Evidence Act and the FSRDC charter explicitly include the goal of supporting evidence-building activity; access by consultants and contractors could expand in the future to provide greater support for such applied research. The FSRDCs have also made it possible for academic researchers to collaborate in many internal Census projects using linked administrative, census, and survey data, including projects developing new privacy protection methods, improving record linkage methods, and developing new federal data products.

3. Which frameworks, policies, practices, or methods show promise in overcoming challenges experienced by governments in their evidence building?

As noted under question 1, challenges experienced by government agencies in their evidence building include a limited pool of skilled researchers within the agencies, a reluctance to make their data available to outside researchers, and the need to combine data from multiple agencies.

For over a quarter-century the FSRDCs have supported and created connections among thousands of researchers from well over a hundred academic and research institutions who collaborate on projects through the shared computing environment. The FSRDCs provide those researchers with secure access to confidential data under procedures that protect privacy and support the researchers in their work. They have experience in hosting data from multiple agencies and are gaining experience with projects that merge data across agencies. This long track record has generated confidence among statistical agency data providers so that there are now five principle statistical agencies (Census, BLS, BEA, NCHS, and NCSES) participating directly in the FSRDCs and over a dozen agencies who trust its security precautions sufficiently to make their restricted data available in the FSRDCs.

The structure of the FSRDCs provides a local researcher community that helps inculcate researchers into the "culture of confidentiality" necessary to safeguard the data. It also supports researchers through the RDC Administrator and Executive Director, who provide outreach about the research possibilities and guidance in developing the research proposal, working with the

datasets, and preparing research results for disclosure avoidance review. The partnership with the academic institutions also provides millions of dollars in direct support of the FSRDC network, making it possible for the federal statistical agencies to leverage the volunteer efforts of the academic research community.

4. The Commission on Evidence-Based Policymaking recommended the creation of a National Secure Data Service. Do you agree with this recommendation, and if so, what should be the essential features of a National Secure Data Service?

Yes, a National Secure Data Service would provide a valuable tool to assist with evidence-based policymaking. Any NSDS should be Accessible, Secure, Efficient, and Equitable. The FSRDCs provide an example of a system embodying these characteristics. FSRDCs are:

Accessible to a wide range of users since they offer a publicly available portal for applications to use the data, expert technical assistance with the process provided by the RDC Administrators and Executive Directors, and secure physical access at the FSRDCs located throughout the country along with opportunities for virtual access to data where that has been approved by the custodial statistical agency.

Secure in their data access that protects the data subjects, data controllers, and researchers. This includes comprehensive vetting and training to authorize users and a secure Virtual Desktop Infrastructure controlling data access. Protection of the data subjects is accomplished through proposal review, de-identification of research data, and statistical disclosure controls over all results to be published, all of which is done inside the FSRDCs, before any results are released for the researcher to use outside the secure setting.

Efficient as the FSRDC network streamlines data discovery, project brainstorming and scoping through assistance from expert RDC Administrators and Executive Directors with experience in proposal writing, training, and technical assistance. Having a broad network saves researchers time and money, as most RDCs have agreements with local organizations that offer researchers who are affiliated with their institutional or consortium members access free of charge.

Equitable since it's not just for "insiders" with connections to agency staff and located near agency headquarters. FSRDCs are located throughout the country and available to most researchers regardless of discipline or tenure. Researchers using FSRDCs include students, professors of all ranks (assistant, associate, and full), employees of think tanks, agency contractors, and other evidence builders.

One model of an NSDS would include a process whereby the data from multiple sources could be linked to create a dataset for a particular project, but the linked research dataset would only exist for the duration of the project rather than being archived permanently. This could raise concerns with research replicability, if the linked research dataset was no longer available. These concerns could weaken the trust placed in the research results, making them less useful to support policy decisions. One possible approach would be to preserve the actual record linkages (without the accompanying data) in a secure but accessible archive through a mechanism such as

the FSRDCs. In that case, future researchers could reconstruct the linked research dataset using the archived linkages, enabling the earlier analysis to be replicated. The FSRDCs could also provide secure access to the linked research dataset while the analysis is conducted.

There are considerable economies of scale made possible through the FSRDCs, as the Administrators gain experience working with researchers on many projects and share that experience with new researchers. The FSRDCs and participating agencies have invested time and effort in developing the procedures and processes that enable large numbers of researchers to work through project development, proposal review, project administration, and final output clearance. Having the projects run through the same system also simplifies things for the partner agencies, since the application process can be documented and routinized, rather than dealing with idiosyncratic requests for data access from individual researchers.

While there are economies of scale in the FSRDCs, the operation is not free. The 32 RDC locations send over $4 million per year to Census, mostly to cover the salary of the RDC Administrators, but also including about $500,000 to cover some costs of program management and disclosure avoidance review. In addition, they provide the physical space for the RDC, the salary of the RDC Executive Director, and other local costs, bringing the total costs to the RDC partners to about $6 million per year. The Census Bureau and other partner agencies spend millions more each year to cover the computing and administrative costs for the FSRDCs.

Expanding the FSRDCs with more research projects, more researchers, and more RDC locations would require additional funding for new locations or expansion of existing locations (even with expanded virtual access), more RDC Administrators, and more staff to manage the administration and conduct disclosure avoidance review. These funds are simply not available within the current funding structure. In addition, further investments should be made in developing common application processes across agencies, improving documentation of the available research data, and providing templates to simplify the various stages of the research process. Finally, some partner agencies charge fees to researchers using their data as a way of recovering their costs of participating in the FSRDCs, which can limit access to their data for researchers with less funding.

5. How can federal agencies protect individual and organizational privacy when using data for evidence building? Recommend specific actions the Office of Management and Budget and/or other federal agencies can take when using data for evidence building, as well as suggested changes to federal laws, policies, and procedures.

It is important to have firewalls in place that separate the data used for statistical and evidence-building purposes from the data used for enforcement and policy implementation. Where possible, data should be de-identified before being made available to researchers and analysts. Sensitive, re-identifiable data should be used only in secure settings and for statistical purposes, with only the results of the statistical analysis being released and with safeguards to ensure privacy protection.

FSRDCs provide an example of one way to protect individual and data controller privacy. Researchers are required to describe their methodology and planned outputs in their research proposals, and these proposals are reviewed by agency staff to identify any disclosure concerns. This forces researchers to think about disclosure issues when designing their project and helps ensure that they don't waste time producing results that cannot be released. Research results obtained within the FSRDCs go through disclosure avoidance review before they are released to the researcher for external use.

One roadblock to expanding the evidence-building capacity at the federal level is the variety of requirements for data access across different agencies. The FSRDCs have been working with partner agencies to identify those different requirements, but developing a more consistent set of requirements would involve changes to agency policies and, in some cases, the laws regarding their data. There are also restrictions on combining data from different agencies which have required separate agreements for each project using those linked data. Developing general agreements for data sharing across agencies would be an important step in facilitating evidence-building research projects.

One area that would be challenging for any NSDS is the need for rapid availability of data and analysis to support decision-makers in emergency circumstances such as Covid-19. Typical evidence-building projects may take years to plan and execute, but during 2020 many projects were carried out in weeks or a few months, combining existing government data with rapidly-available private data. The data service might need to have different procedures or at least some sort of priority system to ensure that results can be obtained quickly enough to be useful for the policy decisions needed to address the emergency.

6. If created, how should a data service be structured to best facilitate (1) research and development of secure data access and confidentiality technologies and methods, (2) and agency adoption of those technologies and techniques?

Especially in the initial stages, it would be important to develop the data service in a setting with high levels of data security. The FSRDCs could provide a secure environment in which to compare current data linkage and confidentiality protections with possible alternatives, such as open differential privacy approaches and model-driven methods to set and monitor privacy budgets.

7. Government agencies have argued that secure data access has value because it (1) improves service delivery, (2) improves efficiency (lowers costs), (3) produces metrics for performance measurement, and (4) produces new learnings/insights from the data. Which of these propositions do you agree holds value and why? Do you have examples that demonstrate these benefits? Do you have other examples of the value of secure data access?

The benefits of secure data access described here could arise from both descriptive and model-driven analyses. A descriptive analysis could identify differences in service quality or costs across different providers within a particular program, providing a way of measuring

performance and encouraging under-performing units to improve. A model-driven analysis could seek to understand reasons for performance differences, leading to recommendations for structural changes in the program that could improve overall performance.

There is also the question of the efficiency of the secure data access process itself. A broad-based secure data access system provides considerable economies of scale. For the FSRDCs, individual RDC Administrators gain experience working with researchers on many projects and share that experience with new researchers. The FSRDCs and participating agencies have invested time and effort in developing the procedures and processes that enable large numbers of researchers to work through project development, proposal review, project administration, and final output clearance. Having the projects run through the same system also simplifies things for the partner agencies, since the application process can be documented and routinized, rather than dealing with idiosyncratic requests for data access from individual researchers. The system is not simply the computing platform, as the FSRDCs now provide multiple (in-person and virtual) tiers of access, or the single application portal, but the network of people who share knowledge and improve practices across universities and statistical agencies.

8. What are the most pressing data needs of state and local decision makers and how would making data accessible from federal agencies help meet those needs? To share data, what guarantees do data owners (or data controllers) need regarding privacy, data stewardship, and retention?

Having a federal data system that incorporates state and local data is important for evidence-building research, even for questions originating at the state and local level. Americans are a mobile people, moving across local and state lines, so integrated data infrastructures are critical to understanding the implications of state and local policies, and allow us to benefit from the experimentation of our federal system. A federal system could establish standards for data quality and documentation, with a common framework for organizing the data coming from different states that would allow research to be conducted more efficiently. Differences across states in their policies can provide clear evidence of the impact of those policies, making datasets covering multiple states much more valuable than a single state's data. Given the mobility of the population across state boundaries, the full impact of one state's policies may not be seen if the analysis is restricted to data from that state – and a better understanding of the factors driving interstate mobility would itself be an important contributor to designing optimal state-level policies. Finally, a federal system accessible to researchers from around the country would expand the potential pool of researchers working on any given problem.

For a federal data system to win the confidence of state and local data providers, its security arrangements would need to inspire their confidence. The FSRDCs offer Safe People, Safe Projects, Safe Settings, Safe Data, and Safe Outputs. Before researchers gain access to the FSRDCs, they undergo a security screening. Before projects are approved, they are reviewed by the agency or agencies providing the data. Each RDC is a secure Census facility, with a keycard-controlled access door, security cameras, and a Census employee on-site. No data reside in the RDC itself but on Census Bureau servers, with all work being done through a secure Virtual Desktop Infrastructure (VDI). Virtual access to FSRDCs leverages the same VDI

technology. Before any research results are released outside the RDC they go through a disclosure avoidance review by the agency providing the data.

9. What are the key problems and use cases where collaborative work between federal, state, and local authorities' data analysis can inform decisions? What are key decision support tools? How would greater communication about data and tools benefit expanded evidence building?

Having a federal data system that also incorporated state and local data is important for evidence-building research, even for questions originating at the state and local level. A federal system could establish standards for data quality and documentation, with a common framework for organizing the data coming from different states that would allow research to be conducted more efficiently. Differences across states in their policies can provide clear evidence of the impact of those policies, making datasets covering multiple states much more valuable than a single state's data. In addition, given the mobility of the population across state boundaries, the full impact of one state's policies may not be seen if the analysis is restricted to data from that state – and a better understanding of the factors driving interstate mobility would itself be an important contributor to designing optimal state-level policies. Finally, a federal system accessible to researchers from around the country would expand the potential pool of researchers working on any given problem.

10. What basic public data services are essential for a data service to address existing capacity gaps and needs? What infrastructure or incentives can the federal government create that locals and states cannot?

Having a federal data system that also incorporated state and local data is important for evidence-building research, even for questions originating at the state and local level. A federal system could establish standards for data quality and documentation, with a common framework for organizing the data coming from different states that would allow research to be conducted more efficiently. Differences across states in their policies can provide clear evidence of the impact of those policies, making datasets covering multiple states much more valuable than a single state's data. In addition, given the mobility of the population across state boundaries, the full impact of one state's policies may not be seen if the analysis is restricted to data from that state – and a better understanding of the factors driving interstate mobility would itself be an important contributor to designing optimal state-level policies. Finally, a federal system accessible to researchers from around the country would expand the potential pool of researchers working on any given problem.