Margaret Levenstein, PhD

Director – Inter-university Consortium for Political and Social Research (ICPSR)

An economist, Levenstein first joined ISR's Survey Research Center (SRC) in 2003 as the executive director of the Michigan Census Research Data Center (MCRDC), a joint project with the U.S. Census Bureau. She has taken an active role at ISR, joining the Director's Advisory Committee on Diversity in 2009 and serving as the chair of ISR's Diversity, Equity and Inclusion strategic planning committee and as the liaison to the larger university program.

Additionally, Levenstein is associate chair of the American Economic Association's Committee on the Status of Women in the Economics Profession and past president of the Business History Conference.

Levenstein received a Ph.D. in economics from Yale University and a B.A. from Barnard College, Columbia University. Her research and teaching interests include industrial organization, competition policy, business history, data confidentiality protection, and the improvement of economic statistics.

Commission on Evidence-Based Policymaking
Hearing, Chicago, January 5, 2017

Submission from: Institute for Social Research, University of Michigan, P.O. Box 1248, Ann Arbor, MI 48106-1248. Contact: Margaret Levenstein, Director, Inter-university Consortium for Political and Social Research and Research Professor, Institute for Social Research, MaggieL@umich.edu.

The University of Michigan's Institute for Social Research (ISR) is pleased to present the comments below for consideration as the Commission deliberates to expand access to and the use of government data for the purposes of building evidence-based evaluation of program and policy outcomes while concurrently protecting the privacy and confidentiality of the citizens and organizations studied. There are several general principles that we believe should guide the Commission as it examines opportunities to build and evaluate evidence-based programs and policies with administrative and survey data. These principles are woven throughout our comments to the Commission's questions below and are summarized as follows:

- Confidentiality of individual data and the independence of the federal statistical system must remain paramount. Participation in our federal data programs, whether they collect survey or administrative data, is premised on the promise that individual data will remain confidential and will be used for statistical purposes only. It is never to be used for enforcement purposes or for the benefit of particular commercial or political interests. Confidence in the estimates produced by our federal statistical system requires adherence to these principles at all times. As articulated in the Office of Management and Budget's Statistical Policy Directive No. 1 (2014), it is critical to "Protect the trust of information providers by ensuring the confidentiality and exclusive statistical use of their responses. Maintaining and enhancing the public's trust in a Federal statistical agency's or recognized statistical unit's ability to protect the integrity of the information provided under a pledge of confidentiality is essential for the completeness and accuracy of statistical information as well as the efficiency and burden of its production." This is just as true when administrative data is re-purposed for statistics. Undermining this trust undermines statistical measurement as well as the effectiveness of the programs upon which the statistics are based.
- Two important steps to uphold these principles and assure the independence and reliability of the estimates produced by our federal statistical system include the following:
    a. Data originally generated outside the federal system, either from state and local, commercial, non-profit, social media, web-based, or other programs, should be aggregated outside the federal system. These data can be cleaned and documented, and secure, confidentiality-protecting crosswalks to PII can be created. Data can then be transferred to the federal statistical system for matching to federal data resources. This will preserve respondent and data provider confidentiality and increase confidence in the security of the system. It

also provides a mechanism for state and local civil servants and third-party (e.g., academic researchers) to access the non-federal data in a secure environment without burdening the federal research data system and its supporting security clearance mechanisms. This is the model that has been adopted by the Institute for Research on Innovation in Science (IRIS) in its collaboration with the Census Bureau's Innovation Measurement Initiative. This model leverages the expertise of those outside the federal statistical system to improve, harmonize, and document the non-federal data. This kind of expertise is often lost or reduced when data are moved exclusively inside the federal statistical system.

b. Data generated by federal agencies and programs will almost surely be legally required to stay within the federal firewall. These data should be made available systematically (promptly, with transparent access procedures, and where there is no or limited documentation, with a mechanism for researchers to contribute to the data and documentation). This will harness the energies and expertise of researchers to improve the data resources of the federal statistical system as well as state-of-the-art analyses of policies in order to assure that the inference drawn from evidence is scientifically sound. This also assures that there is competition in program analysis, so that multiple approaches can contribute to the analysis and program evaluation. The Federal Statistical Research Data Centers are an important mechanism for providing researcher access to these data, but, given the significant ongoing hurdles to their use, they should not be the exclusive mode for researcher access.

Collaborations between federal agencies and academic organizations can help to address several challenges in using administrative data for evidence-based policy making. First and foremost, collaboration with multiple, external academic organizations can help to assure data availability without excessive centralization that might compromise Americans' right to privacy and security in the data that is generated by and about them in the course of their interactions with federal and state agencies. Second, collaboration between the federal government and academic organizations can bring to bear the efforts of the large number of faculty and students who would be willing to work to improve administrative data. This is particularly important as administrative data, like other non-design data, require significant investments in cleaning and documentation to assure that the target measurement concept is what is actually being measured. Third, those same academic collaborators bring with them expertise in both measurement and analysis that contributes to the scientific rigor of the policy analysis. Fourth, collaboration with academic and other research organizations increases the likelihood that multiple analytical approaches are considered, avoiding any tendency for monolithic or even self-serving analysis. Finally, this kind of collaboration develops skills, both on the part of the civil servants on whom we rely for producing the critical statistical resources of our country and and on the part of students who are the next generation of scholars and civil servants. This kind of collaboration will produce a generation of students who better understand the challenges of measurement and policy-relevant analysis as a result of their participation in such a collaboration.

For example, the Institute for Research on Innovation and Science (IRIS), a collaboration of approximately 50 universities based at the Institute for Social Research (ISR) at the University of Michigan, provides an excellent model in which universities (mostly state, but also private) have voluntarily chosen to share confidential, proprietary data, including individual identifiers, with a federal statistical agency for the production of new estimates of the impact of national investments in research and development. IRIS has developed the capacity to ingest, harmonize, and de-identify transaction-level data from its member institutions. It uses these data to produce reports back to its members, restricted datasets available to the research community, and datasets that it transmits to the U.S. Census Bureau for linkage to Census data assets. The Census Bureau produces additional estimates and reports from these linked datasets, and makes the linked data available to qualified researchers with approved projects in the network of Federal Statistical Research Data Centers (FSRDCs). This model has been able to achieve the participation of a large number of institutions, systematic access for the research community, and a more valuable research dataset than was the case for an earlier initiative strictly within the federal government. These data provide the basis for an evidence-based evaluation of a wide range of federal and private programs investing in science and academic R&D. IRIS's model leverages the interest and abilities of the research community to analyze these data as well as leveraging the existing data resources of Census Bureau and the computing resources of the FSRDCs.

In another initiative, in this case between the Institute for Employment Research (IAB) of the German Federal Employment Agency (BA) and the University of Michigan's ISR, now expanded to five other U.S. universities and locations in the UK and on the Continent, hundreds of researchers have contributed to the analysis of German labor market reforms through their access to restricted, linked survey and administrative data. These initiatives demonstrate both the feasibility and the value of academic-government collaborations in overcoming the challenges to creating appropriate data infrastructure and harnessing scientific expertise to analyze those data for evidence-based policy evaluation.

State and local governments produce large amounts of administrative data on programs that they implement, whether funded locally or by the federal government. State and local civil servants have important expertise and knowledge about the operations of these programs and, therefore, the meaning of the administrative data generated by them. They often lack the data or statistical scientific expertise or computing environments in which to analyze or link these resources. Partnerships between state and local governments, federal governments, and academic institutions can provide the relevant training while developing data resources of value to all parties for evidence-based program evaluation.

We recommend two complementary approaches to providing security and assuring privacy. One is to take steps to assure that the data are safe. This can be achieved through traditional anonymization methods (aggregation, suppression, swapping, and noise infusion), data encryption, and the creation of fully or partially synthetic data. The other is to take steps to to assure that the researcher who analyzes the data is safe and is working in a safe computing environment. This is achieved through researcher training and credentialing, scientific peer review and pre-registration of research proposals, and the use of secure computing environments, such as virtual or physical research data enclaves. These two approaches are complementary. Many tasks associated with the work of turning administrative datasets into useful analytical datasets, including data cleaning, the production of metadata, and dataset linkages, can only be accomplished with access to identifiable data. This then requires secure computing *and* a system for training and vetting researchers. The UK's Administrative Data Research Service has made steps in the direction of researcher training and credentialing. The European Union's Data Without Boundaries project envisioned a researcher "passport" to facilitate credentialed access across the European statistical agencies. The Sloan Foundation has recently supported the Inter-university Consortium for Social and Political Research (ICPSR) to build on these earlier projects to establish durable researcher credentials for access to confidential data.

Exploiting the potential power of administrative and survey data for evidence-based policy and program evaluation requires that both government and non-government analysts are able to discover appropriate data resources and gain access necessary to analyze these data effectively. To address these needs, an infrastructure characterized by rich metadata about administrative and survey data sources, a secure platform for researchers to analyze datasets held in other locations, and a standardized and broadly accepted system of researcher credentialing must be developed. The existing network of Federal Statistical Research Data Centers provides an important mechanism for non-governmental researchers to contribute their expertise to the challenges faced by the federal statistical system and to the evaluation of programs.

- The adoption of standardized researcher credentialing, accepted by multiple federal statistical agencies, similar agencies in other countries (such as the German IAB and UK Data), and non-governmental providers of confidential data to the research community (ISR, NORC, RTI) can reduce barriers to accessing data by enabling qualified researchers to analyze data through a modality that is appropriate to their level of data-security training and experience. It also assures that access is obtained for legitimate research purposes on an equitable basis.
- In order to make data useful, and the research arising from it replicable, investments should be made in data documentation via well-defined metadata fields, and infrastructure should be built that enable researchers to locate and analyze datasets held in multiple, distinct, secure locations. Community curation, provided by researchers who are invested in understanding the data and enabled with appropriate software, can assist in building this documentation.

The characteristics of a data-sharing infrastructure designed to increase the availability and use of government data for evidenced-based program evaluation include:

- Robust search and browse capabilities that leverage standardized metadata, permitting researchers to discover data and learn about data in depth
- Capacity to facilitate crowdsourcing (active curation) and improvement of metadata to capture and leverage newly acquired knowledge about the data
- Capacity to recognize varying levels of credentials assigned to a researcher ID
- Functionality that enables researchers to analyze datasets held in multiple, distinct, secure locations, that is, a computing backbone that can support secure, multi-party computing.

Two white papers prepared for the Commission on Evidence-Based Policymaking by the Office of Management and Budget, “Using Administrative and Survey Data to Build Evidence” and “Barriers to Using Administrative Data for Evidence Building,” explicitly point to the key challenges. These include statutory prohibitions that hinder access to the data; policy and legal interpretations, which can vary across agencies and federal, state, and local governments; and resource and capacity constraints, specifically the lack of appropriate and reliable infrastructure to address data sharing and access, management and curation of data, and security and privacy concerns.

The Longitudinal Employer-Household Dynamics/Local Employment Dynamics data program is our best example of such a collaboration. LEHD highlights both the enormous potential and the enormous challenges to creating and making use of linked state and federal data. This collaboration has made possible very important new data assets which have revolutionized our understanding of local and internal labor markets, job creation and destruction, and job mobility for workers in different industries, cohorts, and demographic groups. These valuable data remain underutilized because of limitations to access. Providing resources to strengthen state and local government statistical capacity would allow those agencies and their civil servants to participate more effectively in research using these data. Increased capacity within state and local governments would allow these agencies and civil servants to benefit from collaboration with external researchers and reduce their incentives to impede research.

There are benefits and limitations to both the single- and multiple-clearinghouse approaches. Overall we endorse a principle of union catalogs, so that data can be discovered and compared. A single clearinghouse would facilitate the process of finding and gaining access to the data and potentially linking multiple datasets. A clearinghouse would also act as a single point of entry for an analyst searching for appropriate data with which to address his/her question, and one might expect that a single catalog would have the benefit of consistent metadata to assist the researcher in evaluating the options and identifying the most useful source of data. Having a single clearinghouse to more efficient linking of datasets, for example if the clearinghouse functioned as a trusted third party and provided de-identified, linked data to researchers. A single clearinghouse, using appropriate software to track dataset versions would also increase reproducibility of analyses by make it easier for researchers and policymakers to identify a specific instance of a dataset. Given that administrative datasets are updated regularly as new data are generated versioning of data is particularly important for rigorous and reproducible analysis. The most important benefit to a single clearinghouse is that it would reduce the bureaucratic hurdles to analyses that required to access multiple datasets; on the other hand, these hurdles creates checks and balances and privacy protection that can be undermined by centralization.

Multiple clearinghouses, however, would allow for specialization and expertise around particular data sources and/or types (which lends itself to strong user support as well) and the flexibility to respond more efficiently to changes in formats or uses of data in a particular domain. One of the challenges to using administrative data for research and analysis is the lack of accompanying documentation about the fields in the dataset. A series of specialized clearinghouses could begin to address this because domain-specific staff expertise could, over time, be used to create such documentation -- for example, noting when the underlying meaning of a particular field has changed or even simply pointing out that distributions on key variables changed at a specific point in time so that the researcher could do the detective work necessary to figure out why. Having multiple clearinghouses also spreads and develops the capacity necessary in both person-power (tagging, data checking, user support) and hardware/software for storing and disseminating the data across multiple organizations. This decentralization provides robustness to the infrastructure while increasing privacy protections.

An efficient and privacy-protecting solution would be to have integrated data catalogs and multiple clearinghouses, but secure, multi-party computing across clearinghouses and common standards to gain access to data, including:

- common researcher credentials
- peer review and pre-registration of research project proposals
- data use agreements
- required metadata fields

High quality data requires investment in curation. High quality analyses require investment in training researchers and civil servants and providing them with up-to-date computing facilities. Democracy requires that the data be well-protected. All of these require resources that have to be provided by someone.

One model for self-sustaining data access is the consortium model. Most relevant to the types of data discussed here is the Institute for Research on Innovation and Science (IRIS), a project based at the University of Michigan’s Institute for Social Research. Started with funding from external sources, IRIS’s model was to become self-funded by charging institutions (colleges and universities) an annual membership fee. This fee provides the member institution with campus-level reports based on their data, a seat at the table to help prioritize and design future IRIS products and initiatives, and access to de-identified and aggregate IRIS data for researchers on their campus. One benefit to universities is that, although they are required to deposit data about their campuses with IRIS annually, the IRIS staff has automated the ability to produce charts and reports based upon those data (possible because the same information fields are collected from each institution). An organization or government agency that is required to share data and/or provide reports based on those data can find that it is in their interest to pay those with the skills and resources to properly support the data sharing efforts to carry out those tasks rather than reinventing the wheel and building that capacity within each agency. The data center providing a service such as creating reports or demonstrating use of the data within the research community is seen by the data producer as an added benefit. Similarly, the Inter-university Consortium for Political and Social Research (ICPSR) began as a consortium of 22 institutions in 1962 and continues the model with over 760 member institutions today. These institutions pay an annual membership fee in exchange for access to data curated (and tools created) using member funding as well as reduced tuition for students enrolling in the ICPSR Summer Program in Quantitative Methods of Social Research.

It is important to remember, however, that there is no free lunch; if we want better data and better analysis than currently exists, resources will have to be obtained to support this. The value of this research may well provide the basis for self-financing, but it is more likely that such research creates positive externalities without the ability to generate much revenue to support it.

It is critical that administrative and survey data held by government agencies be made available to qualified researchers and institutions for scientific research and evidence-based analysis. Such access provides the policy community with much greater resources for informing policy decisions than if we rely exclusively on government analysis. It also increases the likelihood that there is diversity in the analytical approaches brought to bear on important policy questions.

“Qualified researchers and institutions” should be established through external scientific peer review of proposals and a system of researcher credentialing that creates an incentive for researchers to be good data stewards. Because of the current lack of consistency across agencies in defining these terms, ICPSR is undertaking a project to research, propose, and test recommendations for researcher credentialing, the result of which will be a tiered set of characteristics that describe “qualified researchers and institutions.” These characteristics will stem from those currently employed/accepted by providers of restricted data, in so far as those requirements are related to protecting against disclosure risk (i.e., not requirements put into place to add bureaucracy or additional “hoops” that must be jumped for data access). We anticipate using factors such as whether one has completed requisite training in ethical data use, is employed at an accredited academic institution, has secured federal funding, and proposes a project that is scientifically sound and that requires access to the data in question.

The ability to disseminate data using a variety of modes (providing metadata only, synthetic data, use restricted to a physical or virtual enclave, or encrypted download) also allows for flexibility in determining access. That is, if a researcher does not have accepted credentials, or is not affiliated with an institution with appropriate technical and legal protections for data, a researcher might still be allowed access to de-identified data. More sensitive data can be restricted to access in a virtual or physical data enclave. In other words, the same data may be made available to different researchers under different access modalities based on the characteristics of the researcher and the sponsoring institution.

The integration (linking) of administrative and/or survey data in a clearinghouse without question increases the risk of disclosure of entities within the data; however, the federal statistical community and the research data community have a long history and reputation for protecting confidentiality. This reputation must be maintained and protected by adhering to the Office of Management and Budget's Directive 1 (2014). Policies to maintain these protections and the reputation of and confidence in the statistical agencies of the United States include:

- Providing access only to credentialed analysts with well-articulated research plans and objectives
- Provision by the clearinghouse of disclosure review of output, notes and other materials that are to be taken out of the clearinghouse (secure environment) to prevent unintended disclosure of subjects within the dataset(s)
- Developing and implementing privacy preserving analytical techniques as well as disclosure avoidance techniques such as creating synthetic populations that preserves statistical information

Clearinghouses can and should require researchers or analysts to submit a detailed proposal of the project for which the data are to be used, specifically addressing why the dataset in question is necessary for addressing the research question. Once these are vetted, by scientific peer review, clearinghouse staff and perhaps an external review board representing the data producer and the study population, a conclusion can be drawn on whether the benefits of the research project outweigh potential risks. Other restrictions should be consistent with the factors listed above -- explicitly agreeing to use the data in an ethical manner (and potentially demonstrating completion of training in doing so), restrictions on the computing environment in which the data can be analyzed, agreeing to terms of use, and the like.

There are a number of private and governmental organizations that offer technological options for data sharing and management. Colectica, a Minneapolis-based firm, is an example of a research and development firm specializing in data management, integration services, and Internet technologies for government, academic, and commercial computing; it offers a range of highly specific products and services useful for supporting data sharing and management. They offer tools for working with metadata using a variety of documenting standards (e.g., the Data Documentation Initiative, DDI,). Colectica also has a portal that offers search, browse, visualization, and data management capabilities.

Other projects exist that could offer either the technology or the functionality considerations that would be helpful. One such project is the Sustainable Environment/Actionable Data (SEAD) project, funded by the National Science Foundation and based at the University of Michigan. SEAD provides a collaborative platform for researchers to curate their data as they undertake analyses, so that the documentation is created and captured and can be harvested when the data are shared (i.e., in a clearinghouse). A number of organizations, such as ICPSR, NORC, and the Michigan Center for the Demography of Aging, use technology to create virtual spaces in which researchers can analyze data that have significant disclosure issues. Generally, these spaces require the researcher to log in to a server housed at the data provider, conduct their analyses, and have output vetted before it is released to them. These virtual data enclaves often disable connections to the internet, print functionality, email, and other programs/features that might compromise data security. Lastly, software (e.g., Fedora) exists for creating and managing digital repositories and could be employed by a clearinghouse.

Resources to train civil servants in state, local, and federal agencies to evaluate their own data will also increase their capacity to learn from and absorb the analyses done by others. The implementation of *multiple* randomized control trials could also reduce the inclination to limit data sharing, as analyses can examine the question of which policies or interventions should be supported at scale (not simply whether an individual policy or program should be continued). Building continuous evaluation and improvement, based on progress toward measurable objectives for the relevant population, into policy design provides programs with incentives to collect and analyze data in order to identify potential improvements.

We should also work to develop a culture that highlights the intrinsic benefit that most civil servants, researchers, and the general population receive from having better answer questions about program effectiveness and other social issues. Researchers and civil servants will then be more likely to suggest improvements to data collection (methods and/or content) that would provide more effective analytic data to use in program evaluation. Researchers who are analyzing data are also likely to catch anomalies or potential inaccuracies that might be missed without researcher engagement. Having multiple researchers with multiple perspectives working with the same data will support models that might be more robust than if a single party were solely responsible for producing the evaluations. The research community's embrace of data transparency and replicability may provide reinforcement to governmental agencies to adhere to similar principles. Sharing data among agencies and with researchers increases the return on investments in data creation. It is more efficient use of government resources. It is rarely the case that a single researcher or organization can study everything that can be examined using a given data source. Differences in disciplinary perspectives mean that data collected for one purpose might be seen by another investigator as having value for his/her project that is completely different. Our statistical agencies employ dedicated civil servants who value improvements in the quality of measurement that they produce for our country. Recognition and respect for these values and these individuals will enable them to be more effective and take the steps necessary to continuously improve our data infrastructure.

There are currently significant barriers in accessing and using such data, including challenges in discovering the existence and location of appropriate data, uncertainty about legal infrastructure and processes for providing access to data, lack of documentation of file contents or data provenance. There are also limited resources for analyzing data (e.g., appropriate training for government employees and non-governmental researchers, appropriate computing infrastructure).

Simply using such information is the first step. Organizations may collect data from program participants but not use it in evaluating the effectiveness of the program, may create summary statistics based on the data but not move further into more sophisticated statistical models, and/or may not be aware of existing research that could inform program/policy decisions. The ability to link data sources provides an opportunity to put data about program participation into context in ways that have not been possible before. For example, having information about students' performance for a given school by itself is helpful, but having the ability to link the information to such things as parental earning records and teacher characteristics allows an educational policy analyst to determine which shifts in student performance are likely a result of new policies implemented at the school, characteristics of the school or teachers themselves, or other issues related to outside influences such as food insecurity. Comparing data across similar contexts or programs is helpful in that the similarities and differences between the contexts create quasi-experimental designs, allowing researchers to identify the parts of the program that are most effective and those where improvement might be needed. Making data available to researchers also provides an avenue for dialogue between academics and policymakers that otherwise might not exist.

Program and policy evaluation should be included in program design so that evaluation is based on evidence that is available and analyzable by multiple, even competing, research teams, held to standards of reproducibility so that all parties can learn from evidence as it accrues in the process of program implementation.

Commission chair Abraham, Commission co-chair Haskins, other commissioners and Dr. Martinez, thank you for holding this hearing today and including me in it.

The development of a statistical infrastructure that integrates administrative and program data, as well as commercial and other non-designed data, has enormous potential to provide the basis for improvements in our knowledge and understanding of the impact and effectiveness of alternative policies. This is an extremely important effort, and I commend you for working to build the statistical infrastructure that we need to create an empirical, evidence-based foundation to undergird policy discussions.

I'd like to emphasize today the important ways that collaborations between federal agencies and academic organizations can help to address several challenges in using administrative data for evidence-based policy making.

First and foremost, collaboration with multiple, external academic organizations can help to assure data availability without excessive centralization that might compromise Americans' right to privacy and security in the data that is generated by and about them in the course of their interactions with federal and state agencies. Confidentiality of individual data and the independence of the federal statistical system must remain paramount. Participation in our federal data programs, whether they collect survey or administrative data, is premised on the promise that individual data will remain confidential and will be used for statistical purposes only, and is never used for enforcement purposes or for the benefit of particular commercial or political interests. As articulated in the Office of Management and Budget's Statistical Policy Directive No. 1 (2014), it is critical to "Protect the trust of information providers by ensuring the confidentiality and exclusive statistical use of their responses." This is just as true when considering administrative data rather than the responses of survey participants. Undermining this trust undermines statistical measurement as well as the effectiveness of the programs which the statistics are intended to measure.

Collaboration with academic organizations can help to address this privacy challenge by providing the basis for a network of data resources that can be analyzed jointly, without concentrating data within a single federal agency. As an example, IRIS, the Institute for Research on Innovation and Science, a collaboration of dozens of universities based at the University of Michigan's Institute for Social Research, is aggregating administrative data from those universities into a single data infrastructure. Those data can be shared with and linked to federal data assets, but they are produced and reside outside the federal government. Other academic collaborations are doing the same for administrative data from state and local governments on K-12 education, criminal justice, transportation, and land use.

This model of a networked data infrastructure, based on collaborations between academic organizations and federal statistical agencies, allows us to reap the benefits from using administrative and program data, especially given the emerging capabilities of secure multi-party computing, without the potential threats to privacy that might be associated with a more centralized system.

Second, collaboration between the federal government and academic organizations can bring to bear the efforts of the large number of faculty and students who would be more than willing to work to improve administrative data. This is particularly important as administrative data, like other non-design data, require significant investments in cleaning to assure that the target measurement concept is what is actually being measured. Investments in preservation and documentation are necessary to meet basic scientific standards of reproducibility; crowdsourcing improvements to data and metadata from the academic research community can do at least some of this. The creation of good documentation and metadata is critical to the effectiveness of a networked system, as it allows for the creation of a "union catalog" of data, and coherent analysis of data, even when those data resources are located in different places. Given the very real resource constraints of the federal statistical agencies, effectively leveraging these external resources is critical to the construction and design of a rigorous statistical infrastructure.

Third, those same academic collaborators bring with them expertise in both measurement and analysis that contribute directly to the scientific rigor of the policy analysis.

Fourth, collaboration with academic and other research organizations increases the likelihood that multiple analytical approaches are considered when evaluating a policy or program, avoiding any tendency for monolithic or even self-serving analysis. Open access to alternative, competing approaches to analytical questions provides the basis for legitimacy of the analysis that is done with administrative data and increases the public's trust and willingness to have data about them and their activities used in this way.

Finally, this kind of collaboration develops skills, both on the part of the civil servants on whom we rely for producing the critical statistical resources of our country and on the part of students who are the next generation of scholars and civil servants. The civil servants who make up the federal statistical system deserve our respect and appreciation. We need to elevate their status, and we do that, not by isolating them within the federal bureaucracy, but by allowing them to engage with their academic peers. This kind of collaboration will also produce a generation of students who better understand the challenges of measurement and policy-relevant analysis as a result of their participation in such a collaboration.

Thank you for your attention and consideration, and thank you for the work that you are doing to modernize our statistical infrastructure, the very basis of our knowledge of ourselves and our society.