

**Commission on Evidence-Based Policymaking:  
December 12, 2016**

**Open Meeting**

**Commissioners Present:**

Katharine G. Abraham, Chair  
Ron Haskins, Co-Chair  
Sherry Glied  
Robert Groves  
Robert W. Hahn  
Hilary Hoynes  
Jeffrey Liebman (via phone)  
Bruce D. Meyer  
Paul Ohm  
Allison B. Orris  
Kathleen Rice  
Kenneth Troske  
Kim Wallin

The open meeting was called to order at 8:30 AM. Chair Katharine Abraham provided an introduction.

**Presentation 1: Census Bureau Models for Acquiring Data, Linking Data, and Making Data Available to Researchers**

*John Abowd and Ron Jarmin, accompanied by Amy O'Hara, Census:*

- Mr. Jarmin noted that the Census Bureau is involved in all aspects of the data lifecycle: acquiring, ingesting, curating, linking, analyzing, and archiving data. Their goal is to increase awareness and acceptance of administrative data in Federal statistics by using such data and developing products that demonstrate their value. Congress, through Title 13, has directed Census to use existing administrative data to support their work and, when possible, help limit survey burden.
- He provided three examples of relevant Census activities: the Center for Administrative Records Research and Applications (CARRA), the Longitudinal-Employer Household Dynamics (LEHD) program, and the Federal Statistical Research Data Centers (FSRDC).
- CARRA includes Federal data, state data, and third-party data. Mr. Jarmin discussed joint projects with several Federal agencies, longitudinal linkage projects with research universities, and evidence-building projects with, for example, Chapin Hall at the University of Chicago.
- Mr. Jarmin discussed the history of the FSRDCs program and its institutional and Federal partners.
- Mr. Jarmin discussed the LEHD, which includes data on firms, jobs, and people from various state and federal data sources. Outside researchers can partner with Census Bureau

researchers to work on internal projects or can work on their own projects through the FSRDCs. Researchers working on external projects have to get approval from each state to use the data.

- Finally, Mr. Jarmin discussed enhancements to infrastructure. They are exploring a cloud-based data facility in partnership with researchers at New York University. They are also working to improve their data management and curation infrastructure and are also pursuing hardware and software upgrades.

### **Discussion and Questions:**

- Chair Abraham asked for an example of a joint project. Ms. O'Hara spoke about a joint statistical project with the Social Security Administration that allows SSA to better understand the demographic characteristics of individuals receiving benefits.
- Chair Abraham asked if all projects must benefit Census. Ms. O'Hara responded that benefits to Census include understanding data quality. Mr. Jarmin added that these projects also help Census develop new data projects.
- Commissioner Groves asked the speakers to comment about the staff skill base needed for an entity that would do similar things. Mr. Jarmin responded that Census has been training staff on big data integration and analysis and has started working with academic institutions to include those topics in their training programs.
- Commissioner Hoynes asked about use of administrative data for the decennial census. Mr. Jarmin responded that CARRA was originally formed to see how administrative data could support the decennial census. In 2020, they plan to use administrative records to assist in non-response follow up.
- Commissioner Troske mentioned that pre-processing may affect the quality of analysis results because researchers may have made different choices during that processing. Mr. Jarmin responded that more collaboration with researchers would be a good thing. Mr. Abowd added that they currently have active projects in which researchers are doing quality reviews of record linkage methodologies.
- Commissioner Orris asked about how the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) has helped with data sharing. Mr. Jarmin responded that CIPSEA allows three statistical agencies to share their business data.
- Commissioner Meyer asked about the barriers to expanding coordination among statistical agencies. Mr. Jarmin mentioned the different rules that apply to data from different agencies. Ms. O'Hara added that governance structures and appropriate information technology resources need thoughtful attention.
- Commissioner Ohm asked about what mechanisms are in place to detect misuse of data. Mr. Jarmin responded that the FSRDC environment is quite secure. Researchers are monitored and all output that leaves the facility is reviewed for disclosure risk. He is unaware of any instance in which confidential data has leaked out of an FSRDC.
- Commissioner Hoynes asked about the adequacy of staffing. Mr. Jarmin responded that just about everything Census does is understaffed.
- Commissioner Wallin asked about metadata. Ms. O'Hara said that they recently developed a data linkage infrastructure website which now has the most basic metadata—a list of

available files. They are working to produce consistent, high-quality metadata for all of the files. They anticipate this will improve researcher proposals to use the data.

15 minute break at 9:45 AM.

## **Presentation 2: Internal Revenue Service Joint Statistical Research Program**

*Barry Johnson, Director, IRS SOI Division:*

- Mr. Johnson provided background on the Statistics of Income (SOI) Division at IRS. SOI uses administrative data exclusively and draw samples from those data for their statistics.
- The Joint Statistical Research Program facilitates access to Federal Tax Information (FTI) for researchers in support of tax administration. Researchers can access administrative data and/or SOI samples. The program is now exploring letting researchers access FTI data through FSRDCs and will be piloting this in 2017.
- SOI allows access to FTI through Interagency Personnel Agreements (IPAs) or student volunteer agreements. A member of SOI staff is assigned to each research project. The staff member then works with the researchers, provides subject matter expertise, reviews all products, and manages the administrative processes at IRS. The staff reductions at IRS over past several years have led to reduced capacity for work with researchers.

### **Discussion and Questions:**

- Chair Abraham mentioned that there seems to be a lot of interest in FTI and asked about the constraints to considering and approving more applications. Mr. Johnson said that constraints include the researchers' own timelines. Graduate students, for example, want to finish a project in a year but the SOI administrative processes take at least 9 months. He further indicated that the biggest constraint is IRS capacity including procurement processes, IT systems, and space. Another constraint is the current legislative requirements for use of administrative tax data.
- Chair Abraham asked about authority to set up a cost recovery system. Mr. Johnson responded that SOI can charge fees to researchers but those do not necessarily come back to the program and cannot be used to support salaries and expenses of staff.
- Co-Chair Haskins asked whether IRS has the desire to accommodate more research, notwithstanding those constraints. Mr. Johnson replied that there is a strong will to continue to offer and even expand researcher access to the data.
- Commissioner Troske asked about SOI's involvement with the FSRDC system. Mr. Johnson replied that they are actively involved in FSRDC governance group and will be piloting access to FTI at the FSRDCs later this year. Mr. Johnson added that a limiting issue is that many researchers want to use the original administrative data rather than the statistical sample. Those data files are large and exceed what is practical to physically move to Census.
- Commissioner Hoynes asked about major impediments to making more use of SOI data as a core backbone for evidence-building. Mr. Johnson said the primary limitation is concerns about the effect of increased use of tax data for research on the voluntary compliance rate in the US tax system. There is a strong belief that voluntary compliance is tied to the strong tradition of keeping such information private. He also mentioned that the cost to curate and manage FTI data in a data warehouse environment is a factor.

- Commissioner Shea asked whether there have been any inappropriate disclosures. Mr. Johnson said no, and that researchers are very conscious of the consequences of a data breach.

**Presentation 3: American Opportunity Study:** *David Grusky, Professor of Sociology, Stanford University and Committee Member, NAS Standing Committee on Creating the AOS*

- Mr. Grusky spoke about the need to develop the capacity to use data to understand economic mobility and monitor population progress towards the American Dream. To develop that capacity, he argues that the U.S. needs a large longitudinal panel.
- Mr. Grusky said that the development of this panel involves three steps: 1) build the population frame using the decennial census and the American Community Survey and, if possible, add tax data; 2) supplement the population frame with administrative data; and 3) link surveys with common identifiers. He added that if tax data could not be linked there could be two frames, one based on decennial census data and one based on tax records.
- To make development of a panel possible, Mr. Grusky said that researchers need to convert earlier decennial censuses to digital format. They are using machine learning processes to do this.
- Mr. Grusky suggested that this should be a major public infrastructure project because there is so much at stake. The potential payoffs include 1) improved monitoring of basic social indicators (poverty, mobility, educational attainment, income and earnings, inequality) and 2) improved program evaluation that assesses intergenerational effects.

**Discussion and Questions:**

- Commissioner Ohm asked about how much loosening of access to these data would be necessary to achieve these goals and who would be prepared to use them. Mr. Grusky responded that there are a number of qualified researchers who are ready and able to use the data effectively to evaluate programs.
- Commissioner Groves asked about the use of Personal Identification Keys (PIKs) to facilitate linking and make it difficult to re-identify individuals. Is the re-identification risk greater when you have 40 years of data on a person? Mr. Grusky indicated that an answer to the question is outside of his area of expertise.
- Co-Chair Haskins asked what the Commission could do to help move the project forward. Mr. Grusky replied that the biggest challenge right now is negotiating data sharing agreements. He further indicated that it would help if the Commission could help ramp up efforts already underway at CARRA and supported efforts to share data from the state level.

**Presentation 4: Bureau of Labor Statistics—Federal State Cooperative:** *Erica Groshen, Commissioner, U.S. Bureau of Labor Statistics, joined by Michael Horrigan from BLS*

- Ms. Groshen described the Quarterly Census of Employment and Wages (QCEW) program at BLS. This program started in 1935 and is a census of establishments that covers 96.5 percent of employees in the U.S.
- Ms. Groshen indicated that QCEW data are acquired through a Federal-state cooperative. BLS provides funding, training, IT systems, and data standards to states, which receive over half of the \$58 million budget. Recent budget cuts have resulted in a slight reduction in data quality. The program has a governance council that sets policies and processes.

- BLS works with states to ensure gold-standard data. BLS staff then add enhancements including linking microdata for each business across quarters to create linked longitudinal microdata. The linked data serves as sampling frame for BLS establishment surveys.
- QCEW has a strong user base that includes Federal offices, outside researchers, and the private sector.
- BLS currently has a program in which researchers can come on site to use the QCEW data, and plan to soon provide access through the FSRDCs. In a recent AEA survey, researchers indicated QCEW was among the top datasets to which they wanted access for research.
- According to Ms. Groshen, QCEW could expand in several ways with little or no respondent burden or cost. For example, QCEW could be linked to information on nonprofit businesses. There is currently a project underway to augment the QCEW with unemployment wage records.

### **Discussion and Questions:**

- Chair Abraham indicated that she was interested in QCEW because it is a Federal-state partnership. There are many other administrative data from programs funded by the Federal government and administered by states. Over time BLS has worked with states to improve data quality. Chair Abraham asked how the relationship between BLS and the states has worked and whether there are lessons to be learned. Ms. Groshen replied that QCEW has both strengths and weaknesses. Its strength comes in the form of input from state partners. The challenge is working with 50 different entities.
- Chair Abraham further asked how important is it that BLS pays states to do this work on QCEW. Ms. Groshen replied that it is essential. Mr. Horrigan added that states have ownership of many of these data, so they have to get permission from each state to share them. He said that it is a natural bureaucratic hurdle to convince 50 states to share their data.
- Commissioner Troske asked why the states have the right to withhold permission and whether there are opportunities for the Federal government to encourage state data sharing. Ms. Groshen said BLS is working on a pilot program to figure out what information states need and build that into new data products. Mr. Horrigan added that states want to share wage records with each other, but it's difficult because states need to reach agreement with the other states. For example, states would benefit from following education of graduates and training completers across state lines for WIOA accountability.
- Commissioner Meyer asked whether the relationship with states has evolved or deteriorated over time and whether personnel changes at state level matter. Ms. Groshen answered that a funding cut at BLS led to a decrease in funding to states for QCEW. That is a problem because it destroys trust. But they have built a robust governance structure that has helped to improve the relationship.
- Commissioner Groves asked for more information about the legal agreements with the states. Ms. Groshen replied that they have 53 different Memoranda of Understanding (MOU), but they try to make them as consistent as possible. Each agreement can only last for one year because BLS only receives one-year funding. Because the government budget process happens late in the year, this has to be a very efficient process. BLS is hoping to move to two-year funding soon.
- Commissioner Rice asked about the extent to which the state agreements are dependent on state law. Ms. Groshen responded that they try to use boiler plate language whenever

possible. Mr. Horrigan added that the pilot to add wage records will require new MOUs with six states.

- Commissioner Groves asked whether states are changing their view on privacy protections for the data they hold. Ms. Groshen said that BLS itself has gotten more restrictive. Co-Chair Haskins asked how. Ms. Groshen said that BLS concerns about privacy are affecting their interpretation of existing laws.

15 minute break at 11:05 AM.

### **Panel: Making Federal Data Available to Researchers**

#### **Presentation 5: Data licensing at the National Center for Education Statistics:**

*Marilyn Seastrom, Chief Statistician, NCES*

- Ms. Seastrom said that NCES has the third largest budget among Federal statistical agencies, but is near the bottom in its number of staff. She provided information about the legal frameworks under which NCES operates. The Family Educational Rights and Privacy Act (FERPA) applies to administrative data from schools or institutions. The Education Sciences Reform Act (ESRA) of 2002 provides for confidentiality of personally identifiable information and says that confidential NCES data are immune from legal processes.
- The NCES licensing system allows qualified external researchers to access data. ESRA allows “agents”, or temporary staff, to access the data. The license agreement is among the user, their organization, and the Institute of Education Sciences (IES, within which NCES is housed). Since the agreements in the license are based on U.S. law, they cannot be enforced outside of the U.S.
- Each license can include up to seven authorized users. NCES can ingest data and make it available to others. Some examples include the Federal Civil Rights Data Collection, state testing data, and graduation rate data. A small proportion of licenses are for use of these linked data.
- NCES licensees agree to unannounced inspections. Licensees are inspected at least once in a project of 1 to 5 years in length. Licensees must submit all work products for disclosure review. NCES aims for a 5-day turnaround for disclosure review. This work is managed by one full-time-equivalent staff person and one full-time-equivalent contractors.

#### **Presentation 6: Research Data Center and Virtual Research Data Center at the Centers for Medicare and Medicaid Services:** *Niall Brennan, Chief Data Officer, CMS*

- Mr. Brennan indicated that CMS is a niche player in the health care system, but that CMS is committed to making data available.
- CMS is embracing a new suite of data and information products, including highly controlled access and open data. Mr. Brennan recommended that Commissioners visit the CMS data navigator at [data.cms.gov](http://data.cms.gov). CMS has created new public use data files since a legal injunction prohibiting the release of identifiable physician information was overturned in 2013. This file contains identifiable information on physicians and hospitals and aggregate data on beneficiaries. CMS has now released dashboards so consumers can use these publicly available data (data for beneficiaries is rolled up to aggregates).
- CMS has established a Research Data Assistance Center (ResDAC) to steer researchers through application process. The Chronic Condition Warehouse (CCW) is the CMS research data warehouse.

- CMS is permitted by law to charge fees to provide data to researchers.
- According to Mr. Brennan, CMS's Virtual Research Data Center (VRDC) is a secure and efficient means for external researchers to access files. the VRDC has helped to expand the user base for CMS data while maintaining confidentiality.

#### **Discussion and Questions:**

- Commissioner Shea asked both speakers if there have been any inappropriate releases. Mr. Brennan said none. Ms. Seastrom said that while there have been violations of the license agreement there has never been a disclosure of individually identifiable information.
- Commissioner Hoynes expressed curiosity about why each organization chose a different access path than the FSRDC system. Mr. Brennan indicated that CMS began by encrypting data files and sending the files directly to researchers. After 20 or 30 years, CMS began using the VDRC model.
- Commissioner Meyer asked Mr. Brennan about why eligibility and claims data for Medicaid from the 1990s are no longer available. Mr. Brennan said that they are available through the ResDAC and CCW but sometimes archive data files when user demand is low.
- Commissioner Orris asked Ms. Seastrom about how the student unit record ban affects the NCES licensing system. Ms. Seastrom indicated that there are types of requests for data they cannot fill. She said that the law prohibiting a national student unit record system at the postsecondary level was extended to cover K-12 at about the same time that Congress provided funding for state unit record systems on education (State Longitudinal Data Systems, or SLDS).
- Commissioner Groves asked Mr. Brennan to elaborate on the CMS fee structure. Mr. Brennan indicated that the prices for accessing data are public and posted on the CMS website. CMS uses the fees to cover government staff time and contractor time to address the requests. The fees help CMS recover some of the costs.
- Chair Abraham asked if CMS can use the fees to hire permanent staff. Mr. Brennan says CMS has not done that yet and he is not sure if they can.
- Commissioner Glied asked about the effect of the Health Insurance Portability and Accountability Act (HIPAA) on CMS data use. Mr. Brennan responded that CMS takes its relationship with HIPAA very seriously. He further said that the requirements of HIPAA are much misunderstood. It is a permissive framework with appropriate safeguards for when data can and should be shared.
- Commissioner Hoynes asked whether there is a disclosure review process at CMS. Mr. Brennan answered that the CMS privacy board serves as their internal disclosure review board.

Chair Abraham **adjourned** the meeting at 12 noon.