

Commission on Evidence-Based Policymaking
January 13, 2017
Open Meeting

Commissioners Present:

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

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

Panel 1: International Models for Acquiring, Linking, and Making Data Available to Researchers and Evaluators

Presentation #1: *Stefan Bender, Deutsche Bundesbank*

- Mr. Bender provided an overview of the German Bundesbank Research Data and Service Centre, which works with the German Federal Statistical Office to develop a broad spectrum of user-oriented economic data. The Bundesbank, like other central banks, produces datasets which are valuable for the production of aggregate statistics.
- Access to data for research in Germany began in 1998 with a bottom-up initiative led by the research community and data producers. The first research data center was established in 2001. As of 2016, there are now 30 accredited Research Data Centers that are overseen by the German Data Forum.
- The Bundesbank Research Data and Service Center (RDSC) is part of an internal project: Integrated MicroData-based Information and Analysis System (IMIDIAS), which has the goal of promoting evidence-based policymaking. The RDSC started in 2014 as part of the Statistics Department of the Bundesbank, and offers access for non-commercial research to highly sensitive microdata of the Bundesbank for free.
- Key principles of the RDSC include: data as a public good, transparent data access, and data protection. In 2016, the RDSC supported 14 employees and over 100 new projects had been initiated.
- Lessons from the development of the microdata initiatives include: 1) development of the access program was fast, but incremental, 2) development required new skills for researchers

and data producers, 3) engagement of researchers was important because of the complexity of the microdata, 4) efficiency and harmonization continue to be a work in progress.

Presentation #2: *Tanvi Desai, Administrative Data Research Network, UK*

- Ms. Desai provided an overview of the Administrative Data Research Network (ADRN), which is a UK-wide partnership providing researcher access to de-identified, linked administrative data in a secure environment.
- When the ADRN was being established, there were concerns expressed around the notion of the development of a population database, and there was negative press around misinformation about data sharing in the UK. The ADRN consciously developed a public engagement strategy. As part of that strategy, ADRN engaged lay members in their governance structure through a “citizens’ panel,” which participates in the review and approval of all research proposals.
- ADRN has a stepwise approach to data linking that includes the participation of a trusted third party organization and maintains a strict separation of activities. Individuals who are actually linking the data never see all of the microdata, and the researchers themselves never see the direct PII. Varied methodology for linking is used; different methodologies depending on the data being matched, the legislation around the data, and the expertise of the researchers doing the linking.
- Approved researchers can only access approved data within a secure setting, such as a physical or virtual data enclave. ADRN has developed “SafePods” that provide a physically secured “pod” that can be delivered to the researcher’s institutional setting.

Presentation #3: *Ivan Thaulow, Statistics Denmark*

- Mr. Thaulow provided an overview of Statistics Denmark, which is the Danish national statistical office. Every person in Denmark has a personal ID, every enterprise has an enterprise ID, and every dwelling has a dwelling ID number.
- The Danish statistical system is still using questionnaires, but to a very limited extent, because most of the data that they receive are generated from the administrative datasets, or registers, maintained by Statistics Denmark, including tax information and data from almost every sector (health, education, etc.).
- Statistics Denmark holds a broad range of data, including: demographic data (since 1980), health data (hospitalizations & diagnoses since 1977; prescription medication since 1995; health insurance since 1992), income (1980), and education (1981). These data comprise the entire population and date back almost 30 years.
- A remote data protocol provides authorized researchers access to data in their own computing environment. Only researchers at authorized institutions can access the data; private companies can be authorized and have access if they have a “stable” research environment.
- Rigorous rules around data access, usage and security have been established in order to access the data. Both the researcher and the researcher’s institution are required to enter into legal agreements that outline the terms of the data access. Violations of these agreements can result in the termination of data access for both the researcher and the institution.

Presentation #4: *Roxane Silberman, CASD, France*

- Ms. Silberman provided an overview of France's Secure Data Access Center (CASD), which was founded in 2009 under the auspices of France's public institution of higher education and research, under the Ministry of Economy and Finance. It operates with technical supervision from the National Institute of Statistics and Economic Studies, one of the agencies in the French statistical system.
- Organized access to confidential government microdata took place in a stepwise fashion:
 - Access to official confidential microdata (census and surveys) since 2008
 - Access to administrative data in France since 2014-2016
- Administrative data is decentralized in France and exists under various legal frameworks, but has benefitted from the centralized process and infrastructure set up for accessing statistical microdata.
- The French effort required changing legal frameworks one by one (tax data in 2014, medical administrative data in 2015, extension to all administrative data in 2016).
- Researchers request data access from the Comité du secret statistique (Statistical Secrets Committee), which meets four times a year to approve projects.
- Approved researchers are trained in legal, IT, and safety aspects of data access, at the conclusion of which, researchers receive an access card keyed to their fingerprints. Researchers are provided with an SD-Box, a secure device installed at their institution, which enables the researcher to conduct their work and access linked data within their own computing environment.

Discussion and Questions:

- Chair Abraham asked how the people selected for the governance structure for the ADRN were chosen. Ms. Desai responded that it was initially quite difficult to get a broad pool, but that there is an extensive public involvement in medical research, and they were able to conduct outreach in the UK through this mechanism as well as traditional advertising.
- Co-Chair Haskins noted that the Commission is very concerned about privacy and security and asked whether any of the international models had experienced a security breach, and if they had, how it was handled? Mr. Thaulow noted that there had been breaches in Denmark, where researchers had taken output home accidentally and improperly, but that the risk was low because the results had been subject to disclosure review. Mr. Thaulow noted that these were not breaches under the law, but rather a breach of the rules. Mr. Bender stated that the Bundesbank had experienced only a few minor breaches—none of which were illegal, even though they did violate the rules.
- Commissioner Groves asked about how the Bundesbank discerns what commercial research is, since he noted that they exclude commercial researchers from access. Mr. Bender responded that this assessment is determined by the institution and whether the institution is considered to be a commercial entity. Ms. Silberman responded that in France their primary concern is whether the results will be published or not. Access is allowed for institutions that are commercial in nature, as long as they publish their findings.
- Commissioner Ohm asked whether the data linked in one of their systems for a particular project remains linked in the system after the conclusion of a project. Mr. Thaulow of Denmark responded that they keep linked data for 20 years, Mr. Bender responded that in Germany data will be stored with the results for at least 10 years, and Ms. Silberman stated in France that they keep the data for 5 years, with an option to extend.

- Commissioner Hoynes asked about the notion of the UK system being a custom service, asking whether data agreements need to be negotiated anew for each project to bring data together. Ms. Desai responded that they act on precedent, meaning that while these negotiations are required, they are much quicker when they have been negotiated previously.
- Commissioner Sweeney noted that Denmark had identified roughly 2,000 unique users, and asked the other panelists how many users there were of their respective systems? Ms. Silberman responded that there were approximately 200 current users of the CASD, Ms. Desai responded that the number of ADRN users is on the order of hundreds, and Mr. Bender responded that he was unsure of how many independent users there were.
- Commissioner Rice asked the panel how each addressed public concerns about privacy. Ms. Desai responded that this has been a key consideration for the UK, and that they had a sizable budget allotted for communications. She noted that transparency is key, including a website that discloses approved projects and outreach events at public gathering places. Mr. Bender noted that Germany conducts no special outreach to the public, as people are generally trusting in the institution. Mr. Thaulow noted that in Denmark, the public recognizes the importance of microdata use and trust is high. Ms. Silberman responded that they focus messaging on the notion of data as a public good.
- Commissioner Sweeney asked whether any of the systems allow individuals to know who has seen their data. Both Mr. Bender and Ms. Silberman responded that both Germany and France afforded this opportunity to their citizens.
- Commissioner Orris asked the panelists if they had an estimated cost per project. Mr. Bender noted that he didn't have a per-project cost calculated for Germany. Ms. Silberman responded that CASD is funded through 2019, and charging researchers for access is a challenge in France. Mr. Thaulow responded that Denmark has an annual budget of about \$3 million. The average cost of a project is about \$2,000, but prices have been decreasing as the work can now be done more efficiently than in the past.
- Commissioner Groves asked Ms. Silberman about the legal history of establishing CASD, asking if their decision to take a stepwise approach to establishing their system (first tax, then medical) was a necessary pattern of legislation or if they might do it differently today. Ms. Silberman stated that tackling each legal issue one by one was the right approach.

Public engagement videos:

- The Commission viewed two videos related to public engagement, one from the ADRN and one from New Zealand.

Panel 2: State Models for Acquiring, Linking, and Making Data Available to Researchers and Evaluators

Presenter #1: *David Mancuso, Washington State Department of Social and Health Services*

- Mr. Mancuso provided an overview of the Integrated Client Database (ICDB) that was established by Washington State's Department of Social and Health Services in the 1990's.
- The ICDB holds over a decade of individual data from state agencies, which is integrated and transformed for analytic purposes in the state social and health service environment.
- The Research and Data Analysis Division (RDA) utilizes the data held in the ICDB in a similar manner as external researchers would to conduct research and evaluation on behalf

of state agencies. The RDA makes their analyses publicly available on the web, a portfolio of work that currently includes over 100 reports.

- Mr. Mancuso described several keys to Washington State's success: (1) Senior agency leadership recognizes potential for integrated data to improve service delivery, (2) The office maintains connections between analytic staff and program operation staff, (3) The office supports service delivery systems rather than academic interests, (4) The office maintains a commitment to analytic integrity to build trust among data providers.
- Mr. Mancuso encouraged the Commission to consider the importance of the state agencies as a critical partner for their expertise, to understand the state data systems, to gain access to subject expertise, and to better understand how to resource states to participate.

Presenter #2: *Kenneth Dodge, Duke University*

- Mr. Dodge provided an overview of the North Carolina Education Research Data Center (NCERDC), which is a collaboration between Duke University and the North Carolina State Department of Public Instruction (DPI).
- NCERDC serves as an agent of DPI by making the state's data on public education accessible to researchers for the purpose of improving education policy for the children of North Carolina.
- Since its inception in 2000, NCERDC has provided support for 375 projects—more than half are initiated by researchers outside of NC
- Current data holdings include annual files of students attending public or charter schools in North Carolina, merged longitudinally back to 1993. These data can also be merged with other files at a researcher's request.
- A 4-person board (2 individuals from Duke and 2 individuals from DPI) establishes policy for the Center, ensures data security, and reviews and approves research proposals.
- Mr. Dodge noted that barriers to data sharing can be overcome with the right set of tools, including statute, standards, incentives, and leadership.

Presentation #3: *Mimmo Parisi, National Strategic Planning and Analysis Research Center, Mississippi State University*

- Mr. Parisi provided an overview of the public engagement strategy implemented in the state of Mississippi during the development of their Statewide Longitudinal Data System (SLDS).
- Mr. Parisi noted that development of a new system could start from a technological approach, or with a public engagement approach. Mississippi chose the latter. He identified a common understanding of the value of integrated data among data providers, data managers, and data users as a critical step.
- According to Mr. Parisi, data governance in Mississippi helps ensure the privacy and confidentiality of all data collected, and allows for the development of common rules guided by a understanding of data value.
- LifeTracks is the SLDS built by the state of Mississippi. LifeTracks connects data from early childhood, education, and workforce data, and the database can be used to produce reports and analysis, or to support research studies.
- One of the biggest challenges in Mississippi was finding a way to convey the value of the data collection system to the state legislature. To accomplish this, Mississippi developed a portion of their website devoted to public accountability with state-level aggregate statistics.

According to Mr. Parisi, the key to sustainability has been the ability to demonstrate value of the program by translating information to be useful.

Presentation #4: *Bob Goerge, Chapin Hall, University of Chicago*

- Mr. Goerge provided an overview of the value of state administrative data and the lessons learned by Chapin Hall in accessing and utilizing state administrative data.
- Mr. Goerge stressed the importance of developing a partnership with data providers in order to increase access to data, as they are not obligated to share their data and they are not always willing participants in research and evaluation.
- In order to build partnerships with programs, Mr. Goerge suggested a need (1) to include input and participation of agency staff, (2) to offer agencies the opportunity to review results before they are shared broadly, and to comment or react to the findings.
- Mr. Goerge stressed the value of state data over data submitted *by* states to Federal agencies, as the data are often transformed, de-identified, or sampled when the data are submitted by the states.

Discussion and Questions

- Chair Abraham asked Mr. Goerge what his reaction was to the notion of directing states to submit raw data to Federal agencies, as opposed to the transformed, de-identified or sampled data that they currently submit. Mr. Goerge responded that this would be advantageous, because though it would be more work to clean and prepare the data for analysis, it would take the burden off the states to complete this task.
- Co-Chair Haskins clarified with Mr. Goerge his understanding that requirements for administrative reporting were not good for researchers. Mr. Mancuso indicated that he was not fully in agreement with Mr. Goerge that the less structured reporting requirements would be less burdensome for states.
- Commissioner Groves commented that gaps in the reporting systems likely reflect gaps in state data collection, and perhaps evaluation should focus more at the state level. Mr. Mancuso concurred, and stated that states are likely the best laboratories for research.
- Co-Chair Haskins asked if Mr. Dodge would suggest that researchers be charged access fees. Mr. Dodge noted that relying on fees that researchers can pay causes the availability of research dollars to drive the use of the data, and that if research funding decreases then funding for supported activities would also decrease. Therefore, additional sources of support for the infrastructure would be critical.
- Commissioner Hoynes asked if charging fees should cause concern regarding equity and access, and Mr. Dodge responded affirmatively, noting that if fees were charged, access would be driven by institutions that have money.
- Commissioner Troske asked about the process by which states had approached obtaining public buy-in for their efforts, noting that the Commission had received a lot of comments regarding the importance of a governance structure. Mr. Mancuso responded that in Washington State, they had not received any pushback on their efforts, likely because of their effort to make the work visible to the public. Mr. Parisi also noted that it was important for Mississippi to convince the public that they understood that the privacy elements and data security were important. Mr. Dodge responded that North Carolina had more recently had gotten a little pushback and some inquiries from state legislators.

- Commissioner Groves asked Mr. Mancuso for suggestions for changing the relationship between the Federal and state data initiatives that would make his work easier. Mr. Mancuso responded that in his state environment, there is so much to do just with their own data, keeping up with the transformation of their own data systems within the state, that they are less interested than others might be in analyzing data from other states. Mr. Parisi offered that the Federal government could instruct states to consider research on the front end when they establish administrative data systems. Mr. Goerge noted that there is still a fair amount of confusion at the state level regarding what data they can share, under what conditions, and with whom. Federal agencies themselves are often not clear, and clarifying the rules around data sharing and broadly sharing these rules with the states would be very helpful.

Panel 4: National Vital Statistics System

Presentation #1: *Shawna Webster, National Association for Public Health Statistics and Information Systems*

- Ms. Webster described the National Association for Public Health Statistics and Information Systems (NAPHSIS), a nonprofit association of the 57 vital records jurisdictions in the United States including the 50 states, five territories, the District of Columbia, and New York City. The collection of vital statistics, including births, deaths, and other life events is the responsibility of the states, not the Federal Government, and access to them is regulated by state law. Vital records are kept confidential to protect privacy and prevent identity theft. They are used to monitor birth outcomes and leading causes of death, among other public health indicators.
- There are administrative and public health uses of vital records. Administrative uses for birth records include proof of citizenship, age, and parentage, and applications for government benefits; death records are used to conduct affairs after death and prevent fraud and abuse. Among the public health uses, vital records are the primary source of data for calculating basic statistics on health across the nation and are regularly used for medical and health research.
- Fees for certified copies of birth and death records and other services partially or fully fund the vital records office in each jurisdiction.
- The vital statistics system faces five main challenges:
 1. Financial- The vital records systems in all states are underfunded and they may not have access to their own revenue from services. Electronic reporting systems are expensive and require long-term, sustainable funding. Recent re-engineering of those systems has resulted in more timely and higher quality data, but much of the work was funded by only by the states themselves. Funds provided by the Federal Government are not enough to keep these systems going. Low wages in state agencies contribute to a high turnover rate and lack of technical expertise.
 2. Political- No one notices the vital records system until it's broken. Without strong leadership, advocacy for these programs suffers within the state. Making them better is not a priority relative to other needs.
 3. Legal- State laws and regulations governing how vital records may be used, and who may use them, vary from state to state.
 4. Capacity- Agencies struggle to maintain the human or technical capacity to keep up with the increasing demands for the data in all the various forms people want them.

5. Lack of authority- Vital records offices don't have the authority to compel data providers (physicians, funeral directors, etc.) to use the electronic systems that they have spent a great deal of money to build and implement.
- NAPHSIS has played a critical and successful role in helping states re-engineer their systems and creating data sharing agreements. NAPHSIS operates a national clearinghouse for birth data called EVVE and will shortly roll out a "fact of death" product as well.

Presentation #2: *Charles Rothwell, National Center for Health Statistics*

- The purpose of vital records is primarily administrative and includes civil registration, national security, and analysis of public health. The first two purposes are a state responsibility and thus their priority.
- Registration of vital events is not a Federal activity—the legal authority resides with the states and jurisdictions that operate the vital records offices.
- The Federal Government is required to collect vital records from the jurisdictions but they are not required to provide them. The negotiated data sharing relationship between the jurisdictions and the Federal Government is the oldest continual federal-state data sharing partnership in the United States. The National Center for Health Statistics (NCHS) establishes contracts with the jurisdictions and requires deliverables such as timeliness and quality. In exchange for the 7 million records a year states provide under the contract, NCHS provides some limited funding, training, and technical assistance.
- Analytical products include Data Briefs, National Vital Statistics Reports, and Vital Statistics Rapid Release—quarterly provisional estimates of natality and mortality.
- Access to data is offered through interactive web-based services, full microdata files for public use, customized restricted data with geographic detail, and restricted access files for researchers. All requests to access customized microdata from restricted data files must be reviewed by NAPHSIS and NCHS for allowable purposes, and if approved, the data are provided at no cost. If the request requires controlled access to microdata because of data confidentiality concerns, researchers may apply and pay to access them through the NCHS Research Data Center.

Discussion and Questions

- Commissioner Orris asked about the impact to vital statistics should the Commission propose a clearinghouse that might place data behind a firewall. Mr. Rothwell replied that states would need to agree to such an arrangement and the clearinghouse would need to be able to continue to link the data as quickly as NCHS can. Mr. Rothwell stated that he supported opportunities for linking vital statistics to additional kinds of data and pointed to the work currently ongoing at CARRA and at NCHS itself.
- Commissioner Glied asked about a National Birth Index. Mr. Rothwell replied that if we had such a system we could link birth certificates to other health records and to the national death index. He said that while we're still a long way from that, we shouldn't put anything in the way of allowing that to happen. The Commission should think about what is possible now that could lead to better things in the future.
- Commissioner Groves asked Ms. Webster about the corporate sponsors of NAPHSIS. Ms. Webster responded that it was important to have sponsors who provide funding without strings attached. In addition, the members (jurisdictions) use a variety of electronic vendors

to do their work so partnerships with those vendors promotes consistency and facilitates linkages.

- Commissioner Groves asked Mr. Rothwell if there has ever been a state or a jurisdiction that dropped out of the vital records system. Mr. Rothwell reported that about five years ago almost all the states considered dropping out because of lack of funding. At that time, the Federal Government could only provide the states with about 6 months of funding. The states need a consistent flow of money to support this work. Turnover in state leadership leads to yearly variability in support for vital records. Emily Holubowich from NAPHSIS added that states rely on the fees from the data sharing agreements. She said that due to state budget shortfalls, some of those funds have been siphoned off into the general fund.
- Commissioner Troske noted that states are reporting data in a fairly similar structure, and wondered how that was happened. Ms. Webster responded that it had been a multi-year effort. They developed a data sharing agreement. A core of the membership pushed it through because they knew it had to be done.
- Commissioner Sweeney said that one of the reasons states moved forward with electronic birth certificates was because they would be useful for research and inform policy. Mr. Rothwell said that there have been some papers, but there is still one state that hasn't gone to the new data format after 20 years because the Federal Government didn't pay for the expansion.
- Commissioner Hoynes noted the value of the vital statistics data and said because of these data we know an incredible amount that we didn't know before. She said these data are easy to access and have protections for confidentiality. Ms. Sweeney asked if those analyses were only at the state level. Several panelists said no. Mr. Rothwell said they've always had a national data file. NCHS started with 10 states and an auxiliary data file to fill in the missing information. Now they have complete coverage although it's taken a long time to get there.

Commission Chair Abraham adjourned the meeting at 12:10 PM.