

# **Barriers to accessing state data and approaches to addressing them**

**Robert M. Goerge**

**June 1, 2017**

## **Abstract**

The technical challenges of accessing large administrative datasets are easily addressed with the advances in data security, computational resources, and the Internet. The most vexing barriers are a set of issues related to legal and ethical issues and control of the data by those agencies that generate it. This article describes those issues and promotes the notion that partnerships with the data providers are necessary in order to facilitate access to researchers, both inside and outside government, but also to provide benefits, in the form of evidence, research and information to the data providers themselves. Ultimately, training of all stakeholders around the secure and responsible use of data and appropriate data stewardship is necessary in order to facilitate the increased use of administrative data that is required to develop evidence that will have an impact on services and programs for individuals and families.

## **Introduction**

While it was not always so, it is now generally accepted that administrative data is a valuable resource for rigorous research and evaluation (Card, Chetty, Feldstein, & Saez, 2010; NSF, 2007).<sup>i</sup> Administrative datasets can contain data on the outcomes of program participation and the characteristics of all program participants. It can contain identifying information of both individuals and organizations and data that is sensitive and confidential because characteristics of specific individuals can be known. Because the outcomes and program data are not always in the same dataset, multiple administrative datasets need to be combined or administrative data can be linked to survey data in order for administrative data to be a powerful tool for doing rigorous research at a large scale in a cost-effective manner. The federal government has recognized the potential value of administrative data by asking the Commission on Evidence-based Policymaking to:

*determine the optimal arrangement for which administrative data on Federal programs and tax expenditures, survey data, and related statistical data series may be integrated and made available to facilitate program evaluation, continuous improvement, policy-relevant research, and cost-benefit analyses by qualified researchers and institutions...*<sup>ii</sup>

The creation of the Commission recognizes that the current state of affairs is not optimal. Administrative data is not often easily available to researchers either inside or outside of government as there are few arrangements whereby the data easily flows to researchers and research organizations. Much has been written and said about the barriers to accessing such data from states for research, policy and program purposes (Goerge, 1997; Hotz, Goerge, Balzekas, & Margolin, 1998; United States General Accounting Office, 1992). However, even as the federal

government states that “(t)he increased use of administrative data for statistical purposes can generate a range of benefits,” little has been done to provide a primary collector of such data—state agencies—with the incentives or resources to improve its use for research and evaluation.<sup>iii</sup> The premise of this paper is that in order to address the barriers, states need to be more closely engaged in the analysis of their data for their benefit.

State agencies are not required to provide access to any researcher. These data are collected primarily for operational purposes, be it determining eligibility for programs, tracking benefits paid to program participants, paying providers of services, as well as frontline case management, the management of prisoners, patients, foster children, students and others for whom the state has responsibility. These systems track data on human resources employed by the state and services provided by NGOs. These systems compile information on taxes paid by individuals and businesses. They collect data on births, deaths, disabilities and diseases for public health and other purposes. Using administrative data for research, be it done by analysts inside or outside of government, is low on the list of primary uses. Furthermore, until 30 years ago, researchers in the domains of employment, criminal and juvenile justice, human services, education, early childhood education, and health care used either survey methods or paper record extraction to compile data to conduct their work. Using administrative data for research is still not routine, but it is certain on the increase as states are expected to become data-driven by elected official and the public. Even those government staff trained in research in universities may not be familiar with the notion of using administrative data for research.

There are multiple stakeholders around the topic of improving access to administrative data from state agencies, even if the stakes are not identified as being specifically about access. Each of the stakeholders can impact access either as providers, users, or facilitators of its use. Clearly, as

the controller of the data, state agencies are the primary stakeholder. Researchers and research organizations increasingly value administrative data to conduct research. This paper focuses on the interests of state agency leaders as the primary decision-makers around access and researchers as the primary user of state administrative data. The value proposition that is made in this paper is that aligning the interests of these two stakeholder groups would benefit both. Closer collaboration and greater communication would help address the barriers and modify what is often an adversarial process between the holders and users of the data.

There are clearly others that have a special interest and can be instrumental in improving the use of a specific set of administrative data in a secure and effective manner. As the funder of many programs and the collection of much of the data, the federal government also has an interest in the quality of federal program management and the implementation of policy, including the protection of human subjects. Governors and state agencies other than that of the state agency holding the data also have a stake in that they may need administrative data to evaluate and improve other state programs. Advocates can be both interested in the analysis of the data as it can point to where classes of individuals may require protection under the law. They, in addition to the agencies that collect the data, are also protectors of the privacy of individuals. Class action lawsuits have been both a positive and negative force around the use of administrative data.<sup>iv</sup> Finally, the public has an interest in that individuals can be served by better use of data, but also harmed by breaches or other improper uses of it, which is why data security and responsible use is of the utmost importance.

This paper describes the barriers and risks, but will primarily focus on how state agencies may benefit from increased access of state administrative data to researchers and research organizations to improve the effectiveness of state programs. A less appealing and likely

politically unfeasible alternative, which would benefit the federal government, is to require state and local agencies to provide data to researchers to evaluate federally-funded programs. Listing the multiple strands of public and political pushback on such a requirement is beyond the scope of this article. The focus is on the better option of moving towards optimizing the benefits of increased access for all stakeholders.

While the focus of this paper is on researcher access to data, both inside and outside of government, the issues relative to other stakeholders accessing an agency's data are very similar. The barriers are similar and the potential approaches to addressing them are as well. The similarities and differences in barriers and approaches to addressing them among all stakeholders will also be discussed.

The technical goal that is the context for this paper is the creation and use of curated longitudinal administrative data that can be linked at the individual level across programmatic and outcome data in a secure, legal, and efficient manner. Access to these data should be governed by data sharing agreements (DSA) or memos of understanding (MOU) that describe the conditions under which researchers could access the administrative data. Other data, such as survey data, could be linked to the longitudinal administrative data to enhance both datasets. Ideally, data across states could be linked.

The approaches that we will discuss converge around state agency leadership become a driving force for the increased analysis of their data to improve federal, state, and local policy and programs. The current circumstances around the access of state administrative data, along with a discussion of barriers, is described, followed by a set of approaches, with some examples, that

bring us closer to the goal of accessing the data necessary to improve the knowledge and evidence that the United States requires to improve the well-being of its citizens.

### **The barriers to access**

This section discusses the barriers to accessing administrative data in a context where both the need for the data and the concerns about how the data are protected are increasing. We lay out the barriers as a set of concerns that the key stakeholders have.

It is important to point out that getting permission to accessing the data and the actual access to the data are two different activities and are likely to be the job of two different units of the organization. The capacity of state agencies to produce a dataset varies tremendously. Some can produce one in days if they regularly curate their own data and it fits the need of the researchers. At the other extreme, there may be no one whose role includes the creation of appropriate datasets and identifying that as a priority could mean months of delay.

### *Risk aversion*

Although access to state agency data for external research and other purposes happens, the controllers of the data—state agency leaders—are often either reluctant to provide data to researchers or simply have not made its use a priority. It is reasonable for them to be reluctant because the results can potentially be harmful to the agency or to a particular leader. Although good data sharing agreements can ensure that the state agency has the ability to review interpretations of results and to provide input, even in these cases, a report may negatively impact the funding for a program and, more importantly, vulnerable individuals and families who rely on the program. Often, the reaction to a critical evaluation may not be to improve the

program, but to end it or reduce its funding. (Recently, a major state workforce support program's funding was cut to almost zero because of weak evidence that it was performing well.)

### *Inadequacy of federal datasets compiled from state data*

State agencies do provide data to other agencies, particularly federal agencies, when there are requirements to do so. This compliance with federal regulations also requires significant resources of the state. In theory, these data, held by federal agencies could provide data for research and evaluation, just as federally collected data, such as tax data, has been shown to be an excellent data resource. However, although federal agencies receive these data from states, the data often have significant prohibitions around re-use—using the data for another purpose (e.g. research). They either cannot be used for general research or evaluation purposes due to federal law (National Database on New Hires at HHS) or require permission from states to be used for research or evaluation purposes (Unemployment Insurance wage data at the Census Bureau). Addressing these restrictions, particularly when the federal government has an interest in knowing what programs are working for whom, would be a step forward especially if states benefit from the increased ability to re-use data send to the federal government. However, legislative change, which would be required in many cases, is slow to come and in some cases, there have been steps backward. Recently, the Child Care Development Fund program (HHS) removed the requirement of providing PII (Social Security Numbers) to HHS for parents and children participating in this program.<sup>v</sup> Another example is the Unemployment Insurance Wage data that the Census Bureau receives under the Longitudinal Employer-Household Dynamics program. In most cases, the Census Bureau cannot use the state data for other purposes without permission from the state.

### *Is it legal?*

Some state agency leaders believe that it is not legal for them to share data with researchers, because in a number of programs, it is the case that administrative data has to be used to improve or benefit the administration of the program. This includes research that is conducted with administrative data in TANF, Medicaid, SNAP, and education data. For example, Medicaid data can be used to improve the administration of the Medicaid program. Education data with PII can be used for “conducting studies for, or on behalf of, schools, school districts, or postsecondary institutions.”<sup>vi vii</sup> IRS data can be used for research if it contributes to tax administration. Clearly, whether or not research supports the administration of the program is open to interpretation. The onus is upon either party to justify that the use of administrative data does inform the administration of the program. An evaluation of a program should improve the administration of the program in some way. If determining whether the program is effective or how it is effective does not impact the administration of the program, it would only be because there is a lack of imagination among all the parties.

The most straightforward mechanism for a research organization accessing administrative data is a contract between the state agency and the research organization, which by definition would be consistent with the administration of the specific program. States may also participate in national evaluation efforts where the intention is to provide evidence for the effectiveness of a new program or policy. External researchers are often part of these efforts as national evaluators or as evaluators for specific states. That being said, this access, as well as contracts, are restrictive in that the product is controlled by the state or federal agency and the administrative data can only be used for that specific evaluation purpose.

### *What's the benefit?*

States do provide data to researchers when they see that there is a benefit to agency. In many cases, as mentioned above, that is a requirement, but one that can be narrowly or broadly interpreted. The fact that these data should be used to improve program administration is potentially the basis for overcoming all other barriers, in that it changes the premise under which state agencies and researchers are discussing the reasons provide access to researchers. The requests for data should always include how states would benefit from the research. If researchers cannot describe the benefit for a state providing the data, then perhaps they should not get access.

### *Defining the legal relationship*

As mentioned above, state administrative data sharing requires a data sharing agreement (or MOU). Once a researcher gets to the stage of creating a DSA with a state agency, some of the barriers will have been addressed. Although templates have been developed, they are seldom used and these DSAs are often created from whole cloth. Good DSAs clearly outline the duties and interests of each party, the ability for the data providers to review draft reports and comment on them, clearly provide permission for specific research activities or how permission for each activity happens, have a duration beyond a single project, official contacts for all parties, and specify requirements for data security. Other features may include the authorizing rules, regulations or legislation, detailed descriptions of the data to be shared, and lists of individuals who will access the data. *All* of the components of data sharing agreements can be considered to be barriers to accessing data and using it. There may be contention on any of these particular issues.

Once a data sharing agreement is in place, the path to data access may be clearer, even if the process does not move quickly. Even with a DSA, state agencies can decide that addressing a particular research question is not in their interest and can employ language in the DSA to reject a particular project.

*It comes down to trust*

Currently, the core of the successful state administrative data sharing is trust between the agency and the individual researcher or research organizations. State agency leaders and staff must trust researchers (and other external entities) to protect data from breaches, to not disclose data or preliminary findings to other organizations (researchers, media, advocacy groups) without explicit permission, or to pursue research which is unknown to the agency. As to the conduct of the research itself, they must trust researchers to protect human subjects and follow IRB regulations where applicable, to understand their programs, to understand the data, to facilitate the review of the methods and findings by the state, to employ that most rigorous methods possible, to work with the state to understand the implications of the research, and to work in good faith to improve the programs of the state agency. Ultimately, the lack of trust between state leaders and researchers is the primary barrier, because of the control that state agency leaders have.

*The need for multi-agency or program data*

Since addressing a particular research question often requires the combination or linking of data across agencies, issues of a researcher having data from multiple agencies is another potential issue that state leadership may be concerned about. When data is combined from multiple sources, new data can be created that increases the potential costs of a breach in data security.

For example, if combining school and health data results in having up-to-date data on the address of an individual with a disability, it increases the potential risk if such information is disclosed. Since such breaches are rare, perhaps a concern that is more likely to happen is that multi-program data will make government leaders worry about what will be said about their program participants on topics for which they do not have data.

State agencies tend to be reluctant to share data with each other. Some of their concerns around sharing with other government agencies are similar to those that they would have in sharing data with external researchers. In addition, there are sometimes concerns about other agencies “knowing their business.” The reality is that state agencies are often in competition for scarce resources, that their programs are at risk of being cut, that their staff may be re-assigned, or that their authority over their operations is diminished as a result of information that is externally compiled. The interactions that many state agencies have with the state’s budget or management offices or legislative oversight body is a common example of this. These factors sometimes actually benefit researchers since state agencies would rather share data with a third party than with each other when the benefit is clear or the risk is low. Again, however, it is difficult to disentangle the complex relationships among stakeholders.

### *Can the data be produced?*

The capacity of states to provide access to external organizations is another potential barrier. Even if state leaders are supportive of providing access and a data sharing agreement can be finalized, the ability of a state agency to physically provide data may be limited. Both the technical and time capacity of state agency staff may be limited. Many states do not create datasets that can be easily shared. If any file that contains the data that a researcher needs does

exist, if it is possible for a researcher to receive and work with that file, accepting existing formats is often the most expedient route to an external organization receiving the data. Any special extracts or reformatting of data may either ultimately prevent access or delay it significantly.

### *Documentation*

Similarly, the capacity of researchers to both assist states in providing them with what is needed or to use the data provided is a potential barrier. This lack of capacity is brought about in part by a lack of documentation about what is available. States have sparse, if any, metadata to inform researcher about the contents or the quality of the data. Because the raw data is typically only used by a few agency staff who work in close proximity to each other, all of the knowledge about a particular dataset may be “in their heads” or kept in documentation that is impossible for an outsider to understand. For example, “H0575” is race of the head of the household of the case. The lack documentation or metadata usually results in a conversation where the agency staff asks what the researcher needs and the researcher asks what is available. This leads to additional delay as either the staff need to pull together information, but more often results in the researchers asking many, many questions over a period of time. Inevitably, the first dataset does not meet the specific needs of the researcher, either in terms of its contents or in terms of the researcher’s ability to use the data.

### *Skills of the researcher*

States today more often than not use relational databases to store their data and the structure of these databases are optimized to support the administration of the relevant program. This means that the database is likely highly normalized to promote speedier processing of individual

transactions. This means that the data that a researcher is asking for may exist in dozens, if not hundreds of individual database tables. For example, one table may only contain the race of individuals in a program, while another only contains the gender. This state of affairs requires either the state programmers to de-normalize the data or for the researcher to do this. Again, this is a source of delay in the researcher getting to the analysis phase of his project. The existence of any type of file that the agency is using to calculate statistics is, as mentioned above, often the quickest way to get most of the data that a researcher might need.

### *Cost*

A significant barrier is obtaining the financial resources to process administrative data into research-ready datasets. While it is often said that using administrative data is a less expensive way to obtain data than primary data collection, the extraction, transformation, and curation of data is an expense that is not easily funded through traditional funding mechanisms. A number of foundations have made a commitment of funding such efforts, particularly in university settings, but as standards and best practices are developed, it is clear that sustainability is a challenge.

While the barriers above have been discussed one-by-one above, it is often impossible for an external party to understand the internal decision-making process of an agency that is deciding whether to provide access to a researcher. The reason that is given for rejection may not be the actual reason. This lack of transparency makes it difficult to address barriers one-by-one. This coincides with most state agencies not having clear policies around accessing their data that are publicly available.

The next sections address approaches that are promising, have worked in some places, and have features that could be generalized. We begin with a set of requirements that are necessary for both the state and the researcher to go forward.

### **Addressing data security**

First and foremost, administrative data must be kept secure and protected from unauthorized access. Without the peace of mind that identified data will not be released or “hacked,” state leaders would not take the first step in providing external access. The good news is that the technology exists to keep it secure and restrict access to those that are fully vetted to view and process the data and more and more state agencies are able to assess the researcher’s ability to keep the data secure. Technical and procedural safeguards must be implemented, maintained, updated, and then communicated to the owners of the data in order for data security not to be a barrier.

A few states and private research organizations have curated data for use by researchers. Best practices—policies and procedures that organizations must implement—are forming as federal law such as HIPAA, FERPA and CFR 42 require specific procedures and policies to be in place. This is the topic of another paper, but it is important for state administrators, who increasingly employ chief information security officers, to have confidence in the research organizations’ data security.

### **Opportunities**

The leadership of state agencies clearly have important programmatic and policy challenges which could be addressed by increased analysis of their own data and likely data from other local, state or federal agencies. In a recent needs assessment, state agency leaders expressed the need to link to data outside of their agencies in order to better understand the characteristics of their program participants and the outcomes that they experience (Weigensberg et al., 2015).

It is difficult for them to pursue this when they cannot attract the workforce or financial resources to both acquire the data that they need and curate high quality data and analyze the databases once built. With a few exceptions, state agencies in the health, human services, education and public safety areas are understaffed.

In order to improve researcher access to state and local agency data, the research and academic communities should fill the gaps of both the leadership's needs of more analysis and their lack of the human resources needed to do it. An effort has to be made by researchers to show that their work can address the needs of state agencies and is not just aimed at their peers. The academic community can train the current and future federal, state, and local agency workforce so that, at a minimum, these agencies can partner with researchers to make their data work for them. Ideally, government analysts could do more to meet their leadership's needs for analysis and help them realize the benefit of more analysis.

Working to address state agency need is typically not rewarded in the university setting, because it often does not impact whether research is published or not. Most researchers do not have incentives to maintain relationships whereby they provide benefits to state agencies over a long period of time, because a couple of well-placed peer-reviewed journal articles can be sufficient

for promotion. It may be that the process of securing data from government agencies needs to be given the same credit in academic departments as designing primary data collection efforts.

Research organizations, which have a mission to produce high quality research that impact public policy, depend on the cooperation of operating state agencies and rely on “soft” money, have a greater incentive to maintain relationships with state agencies and provide technical or research assistance over a sustained period of time. The *quid pro quo* is for these research organizations to receive administrative data from state agencies, often on a regular basis, so that the data can be maintained and the quality can be ensured.

A common complaint among state agency leaders is that they cannot get information from researchers quickly *when they need it*. University faculty are not often in a position to provide information quickly as they often rely on student research assistance and are not ready to drop everything to address the need of a state agency leader. Also, again, this activity provides few benefits for the university faculty member. However, research centers, like the Institute for Research on Poverty at the University of Wisconsin, which curates Wisconsin state agency data does have the ability to provide a quick response. Often, a state leader is extremely appreciative of whatever he or she can get when the agency’s staff cannot provide needed information. This is a capacity issue that could to be addressed to improve access to organizations that can provide a quick response.

The fluency of all stakeholders in the use of administrative data should be continually enhanced. Federal statistical agencies require yearly training on how to handle sensitive data, how to address data security, and the legal requirements and penalties associated with their positions.

This, while very important, is just the first step. As the Big Data meme is upon us, the expectations of government leaders to address the use of their own data have increased. While being data-driven has been a goal for years, the informed general public, including advocates and the media, are asking why government is not using its data better.

However, even when a state leader may provide his or her agency's data for study, data from other state agencies may be necessary for the research. There are multiple efforts that clarify and facilitate the linkage of datasets across programs. Perhaps the best example of this is the use of UI wage data to measure the impact of job training or other programs that have increases earnings as a primary outcome, such as the TANF leaver studies at the time of welfare reform (Cancian, Haveman, Meyer, & Wolfe, 2002). Another example is a study including multiple states that looked at the impact of the Great Recession on the use of Unemployment Insurance and SNAP (Finifter & Prell, 2013). Accessing data from multiple programs or agencies to build comprehensive datasets (Integrated Data Systems, or IDS) is an activity that is at the very least on a few agencies agendas in most states. The State Longitudinal Data Systems program of the Department of Education exists in multiple states and is an example of a database-building activity that can benefit researchers inside and outside of government.

### **Strategies for creating infrastructures for increased data sharing**

A number of efforts to share state administrative data with external researchers have existed over the past three decades, although no one effort has been replicated to the extent that it has become a model for access. What can we learn from these efforts to build more generalized best practice around securing external state administrative data access?

### *State government examples*

State government is not a setting that has been traditionally hospitable to researchers, although there are notable exceptions. Washington State Department of Social and Health Services, since welfare reform, Illinois Department of Children and Family Services, New York State Office of Temporary and Disability Assistance, and Florida Department of Education, among others, have (or have had) research units that operate very successfully within the bureaucratic structure. The question is whether this internal success has implications for external data sharing. In these particular cases, the answer to this question is that it has facilitated data sharing to external research organizations in limited ways. Trusted researchers and research organizations have received data regularly to conduct multiple research projects. The Actionable Intelligence for Social Policy project has, in recent years, brought together states that have built integrated data systems internally along with research organizations that have done the same. Within the AISP group, there is variation in how welcoming government agencies are to sharing data with external research organizations.

A state government-based model is the South Carolina Budget Control and Review Board (now the Revenue and Fiscal Affairs Office). After nearly 30 years of building data capacity across state agencies and programs, South Carolina has the capacity to provide datasets to researchers to conduct research. SC state agencies and qualified researchers can purchase data, at cost, from this agency. Many researchers have benefitted from this model.

Illinois has recently begun the State Data Practice, housed in the Illinois Department of Innovation and Technology, that has as its intention to produce analytics to improve services

provided to state program participants across the human services, health, public safety, and employment. It has employed a data scientist and data architect and will have access to an unprecedented set of data within Illinois. The mandate is to work with state agencies to produce actionable results to achieve the goals that the state has to improve the well-being of its residents.

In the past year, a few states have implemented “enterprise memorandum of understanding” or E-MOUs.<sup>viii</sup> Currently, these E-MOUs do not include external parties, although the intention of the state is to do that. These E-MOUs are intended to facilitate data sharing among state agencies. The intention is that the legalities around data sharing have been addressed in the E-MOU and the process that agencies go about to request data from other agencies is streamlined. Given that these E-MOUs are a new phenomenon, it is yet to be seen whether they will be well implemented and work to improve access within state government. The additional hope is that these E-MOUs can be used as a model to improve access to external organizations and researchers.

### *Federal agency examples*

Yet a different model has been implemented by the U.S. Census Bureau over the past two decades (Johnson, Massey, & O’Hara, 2014). Initially, the Census Bureau began collecting administrative data from federal agencies to improve its ability to address coverage issues in the census count of the U.S. population. Currently, the Center for Administrative Records and Research Applications (CARRA) collects state administrative data to both improve its coverage and data quality, but also to support program evaluation. In most cases, state agencies still have some level of control over the administrative data as specified in agreements between the state

agency and the Census Bureau. CARRA researchers and external researchers can access this administrative data for research purposes (Goerge et al., 2009; Meyer & Goerge, 2011).

The Census Bureau CARRA and the IRS allows for federal data across multiple states to be accessed in certain instances (Chetty, Friedman, & Rockoff, 2014). Building good evidence, in part, may require outcomes to be measured across states. If employment, earnings, incarceration, college attendance, or welfare program utilization, among others, are the intended outcomes to be measured for a particular intervention, looking only within a particular state's data holdings may be insufficient to adequately measure the outcomes. The National Student Clearinghouse collects post-secondary data from nearly all such institutions in the United States, but there are few such examples that are readily available to researchers.

Especially for states where there are high population densities at the borders, state agency leaders might be particularly interested in data from multiple states. Also, given inexpensive travel, higher mobility even among poor individuals may result in the need to measure certain outcomes across states. Therefore, having a data infrastructure that brings together datasets across domains for all states is not only an important tool for evidence-building, but would inform state leaders about the true outcomes for their populations without the bias of just looking within the state's own outcome data.

### *University examples*

Chapin Hall at the University of Chicago and the Center for Urban Poverty at Case Western Reserve are two examples of university-based research organization compiling, curating, linking, and analyzing state data for the purposes of research and evaluation. Their data holdings cut

across multiple decades and multiple state agencies. The value in this model is showing how an external party can effectively facilitate the production of evidence and knowledge through innovations that are valued by state agencies and the research community.

### *Embedded researchers*

One potential approach for improving the communication between researchers and state agency leaders is for researchers to work inside state government as either temporary or permanent staff. The federal government use the Intergovernmental Personnel Act to employ researchers in federal agencies to conduct research that is of mutual benefit to the agency and researcher.<sup>ix</sup> State agencies have employed embedded researchers to assist leadership in the management and implementation of research. This type of activity can be a relatively low cost and quick method to address the barriers described above in that the researcher is brought to the data instead of the data having to leave the agency.

### *Clearinghouse*

Finally, the Census Bureau has funded a university consortium, including New York University and the University of Chicago, to create a prototype administrative data research facility (ADRF) to pilot the notion of an administrative data clearinghouse. The ADRF is being built with state of the art technical resources and including tools that would facilitate the creation of high quality metadata and build a learning community around specific datasets. Such an ADRF could provide services to state agencies, as well as researchers, to reduce costs and provide state-of-the-art tools to facilitate access.

## **Final thoughts**

The primary proposition of this paper is that a closer partnership between state agencies and researchers is necessary to increase access to state agency data by researchers for rigorous evaluation and analysis. Skills in both will need to be developed as the collaboration between the two is not a natural one as the interests of each have not overlapped. Other stakeholders, particularly the federal government, will need to facilitate this collaboration.

This collaboration between government and researchers may seem to blur what should be a clear line between program operators and researchers that ensures objective, independent research. Researchers or research organizations closely affiliated with specific programs many have their objectivity or integrity questioned. Close vigilance to ensure objectivity is needed.

A data security standard to certify organizations to perform the functions listed above for research purposes is needed. Currently, the federal government employs the Fedramp standard, which is required for an entity to have the “authority to operate” a data facility that provides access to researchers. Many states employ the HIPAA standards for covered entities (i.e., health care providers, insurers, government agencies) that do not necessarily apply to research organizations. E-MOU is one piece of this, but states require full solutions that have not yet been specified. Is the Fedramp specification the right one for the states to employ? Or, should states buy-in to a service that is built to address the specific requirements and produce the ideal conditions for the production of evidence that states so desperately need?

Funding to states to transform their raw administrative data into high quality data is necessary. Just as funding is provided to states to collect the data and transmit it to the federal government for compliance purposes, funding should be available to states for this purpose also. If states

would be given the flexibility to use their federal funds earmarked for administration for evaluation, this would also increase the activities needed to increase the use of administrative data.

Data must be curated and the quality must be known to the extent it can be.<sup>x</sup> Where and how that is done is an open issue. Is that the role of the state agency that perhaps knows the data best? Is it up to the researcher to address issues of data quality and communicating them? Or, should this be the purview of a third-party—a government or private organization which has the responsibility for ingesting administrative data, developing metadata, managing legal and physical access to the data, facilitating the analysis of data, fostering discovery and learning about the data, checking for disclosure of the output.

The creation of de-identified datasets—to the extent that does not restrict analysis in anyway—is necessary and must actually be an area of statistical research where we need additional research.

The creation of hashed identification numbers or id numbers that are unique to a particular linkage of a number of datasets are strategies that are beginning to be used more. Active management of the risks around the disclosure of PII is a requirement that must be addressed.

Researchers should only have access to data that they require to conduct rigorous research.

Superfluous data should not be accessed.

## **Recommendations**

- Encourage ongoing collaborations among state and local agencies and researchers in order to jointly address the barriers in using administrative data across programs and agencies.
- Build collections of data in secure facilities with the proper controls to ensure that only those individuals with the proper permission have access to data in a quick, manageable fashion.

- Develop and hire agency leadership that understands the need for evaluation and research
- Train state and local government staff in the use of administrative data for program management and evaluation
- Train researchers not only in the techniques necessary to process and analyze administrative data, but also in state information system contents and database technologies that will allow them to facilitate the physical transfer of data from state agencies.

---

<sup>i</sup> Archives of de-identified or open data exist and can lead to important research, but these are limited in their ability to support program evaluation (i.e. Adoption and Foster Care Report System, National Child Abuse and Neglect Data System, or state or municipal data portals).

<sup>ii</sup> Evidence-Based Policy Commission Act of 2016, Pub. L. No 114-140

<sup>iii</sup> OMB 14-06 “Guidance for Providing and Using Administrative Data for Statistical Purposes”

<sup>iv</sup> Administrative records have been used by advocates to document the poor treatment of special populations. Also, legal advocates have sued governments for the illegal use of administrative records. See <http://www.childrensrights.org/our-campaigns/class-actions/>

<sup>v</sup> <https://www.acf.hhs.gov/occ/resource/overview-of-proposed-changes-to-acf-800-and-acf-801-ccdf-administrative-data-reports>

<sup>vi</sup> [http://ptac.ed.gov/sites/default/files/Guidance\\_for\\_Reasonable\\_Methods.pdf](http://ptac.ed.gov/sites/default/files/Guidance_for_Reasonable_Methods.pdf)

<sup>vii</sup> [http://ptac.ed.gov/sites/default/files/FERPA%20Exceptions\\_HANDOUT\\_horizontal\\_0.pdf](http://ptac.ed.gov/sites/default/files/FERPA%20Exceptions_HANDOUT_horizontal_0.pdf)

<sup>viii</sup> Virginia and Illinois are two states that have implemented E-MOUs.

<sup>ix</sup> <https://www.opm.gov/policy-data-oversight/hiring-information/intergovernment-personnel-act/>

<sup>x</sup> “Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle. The active management of research data reduces threats to their long-term research value and mitigates the risk of digital obsolescence.” From <http://www.dcc.ac.uk/digital-curation/what-digital-curation> December 6, 2016.