



THE CONSORTIUM
stimulating self-sufficiency & stability scholarship

THE FAMILY SELF-SUFFICIENCY DATA CENTER: LESSONS LEARNED

SCOTT W. ALLARD* | University of Washington

EMILY R. WIEGAND | Chapin Hall at the University of Chicago

ROBERT M. GOERGE* | Chapin Hall at the University of Chicago;
University of Chicago Harris School of Public Policy

LEAH GJERTSON | Chapin Hall at the University of Chicago

*Co-Director, Family Self-Sufficiency Data Center



 **CHAPIN HALL**
AT THE UNIVERSITY OF CHICAGO



This research was supported by the Family Self-Sufficiency Research Consortium, Grant Number #90PD0272, funded by the Office of Planning, Research, and Evaluation in the Administration for Children and Families, U.S. Department of Health and Human Services. The Family Self-Sufficiency Data Center (FSSDC) facilitates the use of administrative data by researchers and administrators to improve understanding of and identify methods for increasing family well-being. The authors would like to thank all the staff from the University of Chicago Harris School of Public Policy, University of Washington Evans School of Public Policy and Governance, Chapin Hall at the University of Chicago, NORC at the University of Chicago, and Orlin Research who contributed to this project. We are also grateful to our federal partners for their advice and support, especially Nicole Deterding, Emily Schmitt, Naomi Goldstein, and Mark Fucello. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the Office of Planning, Research, and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

Recommended citation:

Allard, S. W., Wiegand, E. R., Goerge, R. M. & Gjertson, L. (2020). *The Family Self-Sufficiency Data Center: Lessons learned*. Washington, DC: Family Self-Sufficiency and Stability Research Consortium.

EXECUTIVE SUMMARY

Today, more than ever before, administrative data play a central role in policy-driven research on poverty and self-sufficiency. Policymakers and government leaders at federal, state, and local levels seek more research and analysis on applied questions, both by analysts within government and by external research partners. Heightened interest in the potential of administrative data and a growing body of research drawn from administrative data sources have been accompanied by the documentation of challenges and recommendations for pursuing this type of work. However, the field needs resources that offer a critical examination of strategies and activities commonly used to encourage effective, rigorous, and ethical administrative data use. Too few of these resources exist now.

To begin to fill this gap, this report lays out lessons learned from the Family Self-Sufficiency Data Center (FSSDC or Data Center) initiative, a 6-year effort to advance the application of administrative data in the human services. The FSSDC was created in 2013 through a grant from the Office of Planning, Research and Evaluation (OPRE) in the Administration for Children and Families (ACF) at the U.S. Department of Health and Human Services (HHS). The Data Center aims to enhance the availability, quality, and use of administrative data for family self-sufficiency research and analysis.

FSSDC efforts to advance the use of administrative data to improve family self-sufficiency fall into four categories: (1) compiling and preparing administrative data sources; (2) technical assistance activities to improve agencies' use of data; (3) developing software, tools, and resources to support data use; and (4) dissemination of research on needs, barriers, and challenges in using administrative data. In each category, we draw from our experiences over six years to assess the impact of different approaches and make recommendations intended to help maximize the impact of future endeavors in this space.

Compiling and Preparing Family Self-Sufficiency Data Sources

A component of the initial vision for the Family Self-Sufficiency

Data Center was to create a secure, electronic repository of family self-sufficiency datasets across programs and jurisdictions that allows state agency and external researchers to access linked administrative data. Despite the promise of this approach, we moved away from it within the first two years of the project due to a number of governance-related challenges. In particular, the FSSDC struggled most with questions around data sharing, data permissions, and creating a data repository environment that provides data security while also accommodating the needs of real-time policy work.

However, as detailed in this report, the FSSDC had some success linking and analyzing data through other channels. We describe a partnership with the HHS Office of Family Assistance (OFA) that compiled and organized federally reported Temporary Assistance for Needy Families (TANF) microdata for the years 2009 to 2017. We developed a longitudinal data file and a report detailing our findings with regard to the data contents, quality, and fitness for various analytic purposes, including some early caseload dynamics analyses.

Lessons learned: Ownership and buy-in from the agency data providers is a key driver of the success or failure of any effort to create a repository. Data providers need to see clear benefits and long-term sustainable operations that justify their time and efforts. Options for developing buy-in to a data repository or comparable initiative include:

- Clear messaging about the purpose and operation of a data repository with articulated benefits for each stakeholder;
- The opportunity to streamline existing, burdensome data sharing and governance activities, to reduce burden on staff and agency resources;
- Assistance with creating and maintaining data documentation needed to put data into a repository; and
- Increased access to data from other programs with relevant information about participants.

Technical Assistance Activities to Improve Agencies' Use of Data

Across the six years of the initiative, the FSSDC experimented with various technical assistance (TA) strategies to support administrative data use. Initially, we provided one-on-one assistance to a small number of pilot states around data quality, linking, and analysis. Those engagements led us to deliver large-format technical assistance for agency staff from a large number of states. These TA efforts involved workshops and hosted events, as well as a LinkedIn group and email newsletter. Later, we explored small group technical assistance, with the idea that states working together in small groups could identify issues that were not unique to any single state and might have general solutions. In that multistate project, we offered a series of interactive technical assistance sessions where we supported a small group of state agency staff working on a similar analytic task.

Lessons learned: Pursuing and experimenting with a range of technical assistance efforts during the six years of the initiative generated many insights into how such support might be provided in the future.

- Consistent staff capacity and motivation are key to successful engagements. Staff turnover and competing urgent priorities often threaten partnerships at very early stages.
- Persistent partnerships can be facilitated by the provision of resources or other short-term incentives and by achieving small wins, in combination with properly managing expectations to help agencies stay invested long-term.

- Even in successful partnerships, moving from improving technical outcomes to impacting the program or participant well-being remains a significant charge. Often, these areas of impact exist in inherently political spaces, where analytics are only one consideration. Such realities should shape expectations around TA.
- States regularly express strong interest in working in groups and in sharing knowledge with their peers, but time and resource constraints on all sides often pose challenges for getting these partnerships beyond the superficial.

Developing Software, Tools, and Resources to Support Data Use

The FSSDC also wanted to develop tools or resources that could be adapted to data users' specific data context and infrastructure. These resources would present modular solutions to address common problems we identified as we worked with partners. Over the six years of the initiative, we developed a series of resources and products. Our first effort, a data tool that would help states access and analyze their administrative data in a secure environment, did not fit well into agency work routines. In turn, we shifted our focus to modular resources. We created a TANF data model modular resource that helps users manipulate administrative data into a usable analytic format. Along with training materials that discuss the logic of the data transformation processes, the FSSDC team developed scripts in both R and Python to provide states with common methodologies that would facilitate caseload analyses. Eventually, the FSSDC built a GitHub repository to serve as a platform to share scripts and other resources with state agencies. Finally, the Data Center team created a Data File Orientation Toolkit, meant for administrative data users seeking to understand aspects of data quality.

Lessons learned: Our work provided insight into how tools and resources might be developed in order to increase the probability of use and adoption.

- It is important to make the learning curve for any new tool or resource as low as possible. Simpler tools or resources that address a single challenge, rather than "one size fits all" solutions, require learning fewer features. Pairing resources with very clear, simple documentation and multiple channels or methods for learning can facilitate adoption. Also, wherever possible, it always helps to build off existing areas of familiarity for the agency and the potential user.
- Role models of successful adoption of a tool in other state agency contexts can be instrumental in fostering additional adoption by providing motivation and concrete examples to follow.
- All the best adoption practices will not matter if the tool doesn't meet a demonstrated need. Software, code, and resource solutions should be carefully crafted to complement the gaps and challenges identified through technical assistance activities.

Continuous Learning about Needs, Barriers, and Challenges around Administrative Data

In our final area of focus, we developed resources that accurately trace current administrative data needs and barriers to use. The FSSDC completed a thorough needs assessment in 2013. This assessment led to a report that examines the administrative data resources available and the realities of how agencies use administrative data to inform policy or practice (Weigensberg et al., 2014). It also led to a subsequent journal article that used findings from the needs assessment to develop roadmaps to greater analytic capacity (Allard et al., 2018). Similar work was completed around state agency capacity, opportunities, and open challenges related to administrative data record linkage (Wiegand & Goerge, 2019c, 2019b, 2019a).

Lessons learned: Several conclusions and recommendations emerged from these learning activities:

- We repeatedly learned how difficult it is to understand what is truly needed to increase data use capacity. There were discrepancies between what people said they needed and what they actually needed. There were also discrepancies between our initial assessments of what was needed in a particular agency environment and what was ultimately needed.
- The presenting challenge to data use may not be immediately identifiable to an agency or particular analyst. Often the reason agencies need customized or one-on-one technical assistance is because they can't pinpoint exactly where things are going wrong.
- Attempts to document barriers to data use and capacity-building needs should be mindful of these challenges. Ask about challenges (in multiple and creative ways, if possible), seek more detailed evidence, and be willing to reconsider initial assumptions.

Overall Conclusions and Recommendations

Across our many different areas of activity, several key themes about pathways to improve the quality and use of administrative data consistently emerged:

- Support from leadership is necessary to make and sustain progress around data use.
- Flexibility and experimentation are crucial. It can be difficult to identify exactly what supports are needed, and those gaps may shift over time.
- Recognize that effective data use is difficult for a variety of complex reasons. Help agencies identify key problems and recognize that those problems are often not unique. Make progress by segmenting challenges into small, manageable areas for intervention and presenting resources to overcome those hurdles.
- Balance the tension between providing customized TA (which can be expensive and difficult to sustain) and generalized TA (difficult to make relevant to each individual circumstance) with strategic application of custom and generalized support and the use of hybrid approaches.
- The technology landscape is shifting, requiring adaptability. Technologies that once created significant barriers, like cloud storage of data or analytics in R or Python, are becoming increasingly commonplace in state government.

We summarize the application of these findings to difference audiences in a table on p.13.

INTRODUCTION

Today, administrative data play a more central role in policy-driven research around poverty and self-sufficiency than ever before. There are three parts making up this trend:

1. The increased expense and sampling challenges of survey data (Krosnick, Presser, Fealing, Ruggles, & Vannette, 2015);
2. A greater recognition of the value of administrative sources (Federal Data Strategy Team, 2019a; Hotz, Goerge, Balzekas, & Margolin, 1998; National Academies of Sciences, Engineering, & Medicine, 2017; Penner & Dodge, 2019); and
3. The growing popularity of data science or "big data" initiatives as approaches to solve vexing social problems (Connelly, Playford, Gayle, & Dibben, 2016; Schroeder, 2014).

In particular, researchers interested in understanding social welfare program dynamics are using, or seeking to use, administrative

data from state and local government agencies. Policymakers and government leaders at federal, state, and local levels seek to encourage more research and analysis on applied questions, both by analysts within government and by external research partners.

Heightened interest in the potential of administrative data and a growing body of policy research using administrative data have been accompanied by documented challenges and recommendations for pursuing this type of work. Existing reports describe known barriers to utilizing administrative data policy research with governmental entities. These reports also share successful examples and recommended strategies for overcoming challenges (see, for example, Doar & Gibbs, 2017; Duran, Wilson, & Carroll, 2005; Federal Data Strategy Team, 2019b; Goerge, 2018; Maxwell, 2017; Office of Planning, Research and Evaluation, 2016). Other papers provide targeted guidance for pursuing work with administrative data through the development of data and research partnerships between government agencies and external researchers (see, for example, Green et al., 2015; Lin, King, Maxwell, & Shaw, 2018; and National Forum on Education Statistics, 2012).

Often absent from the existing literature are in-depth discussions of particular activities that have, and have not, worked to facilitate these efforts with governmental agencies from multiple jurisdictions. Few resources critically examine specific strategies for encouraging effective, rigorous, and ethical administrative data use.

To begin to fill this gap, this report lays out lessons learned from the Family Self-Sufficiency Data Center initiative, a six year effort to advance the application of administrative program data within human service agencies. The effort began with a comprehensive needs assessment. It subsequently included a range of training and technical assistance (TA) activities, including workshops and peer convenings, one-on-one TA, small group TA, and the development of software tools, code, and written resources. Drawing from this wealth of experience, we discuss the opportunities and challenges to each approach and make recommendations to position other, similar initiatives to draw from our experience and maximize impact.

The Family Self-Sufficiency Data Center

The Family Self-Sufficiency Data Center (FSSDC or the Data Center) was created in 2013 through a multi-year grant from the Office of Planning, Research and Evaluation (OPRE) in the Administration for Children and Family (ACF) at the U.S. Department of Health and Human Services (HHS). The FSSDC aims to enhance the availability, quality, and use of administrative data for family self-sufficiency research and analysis. Additionally, the FSSDC seeks to advance understanding around the policy and organizational issues that either facilitate or hamper administrative data use. Through written reports, TA, and conference presentations, the FSSDC shares these understandings with the field. The FSSDC pursues this work through partnership with researchers, policymakers, and administrators to answer fundamental policy and program questions and build knowledge that will be translated into better policy and practice. Of particular importance, the FSSDC was part of the Family Self-Sufficiency Research Consortium (FSSRC), a five year effort that brought together seven family self-sufficiency researchers. This team worked to find evidence-based strategies to improve the lives of low-income families and children through better policies and practices by integrating research, policy, and practice on family self-sufficiency and stability. FSSDC team members met biannually with the FSSRC to share research insights and lessons from applications of administrative data to policy-making settings. Partnership with the FSSRC helped the FSSDC connect with data partners, data resources, and learning communities.

As a six year pilot effort, the FSSDC combined the scholarly leadership of the University of Chicago and the University of Washington, the expertise of Chapin Hall in working with state and local administrative data, and the NORC Data Enclave's data storage, processing, security, and dissemination capabilities to provide additional expertise on data access and management. Together with leadership from OPRE and input from ACF offices, the Data Center supported state agency staff and academic researchers' efforts to promote, use, and disseminate data broadly relating to family self-sufficiency. Given OPRE's policy focus and the FSSRC's emphasis on family self-sufficiency, the FSSDC primarily targeted state-level health and human service agencies.

Our efforts to advance the use of administrative data to improve family self-sufficiency fall into four categories:

1. Compiling and preparing administrative data for analyses related to family self-sufficiency across programs and jurisdictions in a secure, centralized data store;
2. Technical assistance activities, such as workshops, one-on-one, and small group projects;
3. Development of software tools, code, and resources to support researchers and analysts in the preparation and use of family self-sufficiency data; and
4. Writing and dissemination activities to reach the broader field of researchers, advocates, and funders and to articulate barriers, challenges, and needs in improving the use of administrative data in family self-sufficiency.

This report reviews each of these four categories. In each area, we summarize our experiences, including successes and failures, and describe our recommendations for similar activities in the future. Appendix A contains a list of research products created through the Data Center.

COMPILING AND PREPARING FAMILY SELF-SUFFICIENCY DATA SOURCES

Our Experiences

The initial vision for the Family Self-Sufficiency Data Center was to create a literal data center: a secure, electronic repository of family self-sufficiency datasets across programs and jurisdictions. The repository would have data linking and governance procedures to facilitate the use of these data to investigate complex family self-sufficiency questions. Yet, it would also respect confidentiality of program participants and the legal and ethical responsibilities of data providers (i.e., the agencies from which the administrative records originate). Administrative data repositories, which include integrated data systems, allow state agency and external researchers to access linked administrative data stored in secure settings. Repositories are commonly proposed tools for enhancing administrative data use. Such repositories can be found in a small number of state governments (e.g., the Office of Research and Data Analysis in the Washington State Department of Social and Health Services; South Carolina Budget and Control Board, Division of Research and Statistics; and the Virginia Longitudinal Data System). Other initiatives, often privately funded, have built repositories that exist outside of state agencies (e.g., Policy Lab at Brown University; California Policy Lab; Center for Analysis of Longitudinal Data in Educational Research; the Administrative Data Research Facility at NYU).

Yet, there are a number of governance-related challenges to creating data repositories, and we moved away from creating such a repository within the first two years of the project. Ultimately,

we struggled most with questions of data sharing and permissions. Although we had negotiated data sharing agreements (DSAs) with several pilot states, these were solely to allow data center staff to assist with internal needs; we encountered a combination of resistance to, and low motivation for, any broader collaboration that included sharing data. Even negotiating the limited DSAs we executed often proved to be lengthy and time-consuming processes. In our pilot states, in a period of several months to a year, there were informal processes discussing data sharing possibilities. Subsequent formal processes to establish DSAs and access data took an additional 12 to 18 months. Moreover, we found many agency partners did not have staff capacity or continuity necessary to pursue more ambitious applications of administrative data once DSAs were signed.

In general, we found that while researchers saw and felt the immediate potential benefits of a centralized data repository, state agencies were much more hesitant to participate in such efforts. Several states already had pre-existing repositories or data sharing arrangements to which they were committing time and resources. Where a data repository represented a potential new source of data to researchers and agency staff, many state agencies were concerned about increased liability and greater exposure to potential ethical or legal challenges to sharing data. These concerns outweighed any potential benefits. Agencies voiced concerns about lack of control over data use by external research partners. They also expressed frequent concerns about the security of hosting data externally or in the cloud. (Since the start of the FSSDC, however, the landscape and awareness of cloud computing has changed significantly. State agencies better understand its power and agility as well as its security proposition. We do not expect that future efforts to create administrative data repositories would face this particular challenge.)

We found another significant impediment to be a lack of clarity within federal and state regulations defining the circumstances in which administrative data can be shared. Some states said this lack of clarity impeded them as they weighed opportunities to share data internally across agencies or with external researchers. For example, many programs allow data to be shared with external organizations if the purpose of the sharing is to improve the functioning of the program. Some states referenced limitations imposed by federal Health Insurance Portability and Accountability Act (HIPAA) or Family Educational Rights and Privacy Act (FERPA) guidelines, even when it was not clear that those policies governed the data at hand. At other times, federal or state agencies provided limited guidance about what qualified as appropriate use of administrative data, leading to inertia around discussions of data use. For example, state program data related to the Supplemental Nutrition Assistance Program (SNAP) was particularly relevant to the work and mission of the FSSDC and its partners. Yet, the U.S. Department of Agriculture (USDA) provided little clear guidance regarding appropriate use of the program data maintained by states. State program directors consistently reported that they did not have sufficient clarity to make data sharing decisions. In all these instances, ambiguity made state agencies risk-averse and often led them to hedge on sharing data out of concern for potential legal ramifications.

The nature of the technical environment we provided state agencies to mitigate potential security risks further complicated attempts to create a repository across states and for different types of external users. The demands of security and control over data use

can lead to user experiences that are often ill-suited for real-time policy work. For example, disclosure review requirements added time lags and processes to the export of analytic results. Working in a restricted environment required more effort and long-range planning than was feasible in many instances, particularly given many staff had limited time to dedicate to analysis and given the short lead times for inquiries from legislative or agency leadership.

Questions about the sustainability of a new data repository also led agencies to hesitate to dedicate time and effort to creating data sharing agreements. It is common for data repositories to be launched with multiyear grants, with the expectation that subsequent sustainable funding from other sources will be acquired during the grant period. Even though the FSSDC was initially funded for five years, it was understood the planning and administrative work to secure DSAs and transmit data would take several years. As a result, agency partners commonly wondered whether there would be enough time to accomplish meaningful work. Without clearly defined funding for a sustainable future by the second year of our five year initiative, it was not clear whether the repository would persist past the initial start-up endeavor. Therefore, several potential agencies were not inclined to invest time in questions of data governance, in crafting and executing data sharing agreements, and in developing appropriate data extracts and metadata. It is pretty clear that unless funders make a sustained commitment of hard money to the necessary staffing and infrastructure, an initiative to promote use of administrative data will always be a risky proposition for state partners. These issues may be particularly acute when seeking to build a repository that involves multiple state or federal partners. While a time-limited project such as the FSSDC can pilot certain activities, we are convinced that federal and state government must ultimately commit earmarked funds for periods longer than five years. Such long-term funding commitments, however, run contrary to the needs of funders to generate immediate benefits of data sharing and linking.

Building on these experiences, FSSDC staff also participated in the development of the Administrative Data Research Facility (ADRF) at New York University (NYU), which is a FedRAMP-compliant,¹ cloud-based secure platform that hosts confidential microdata. The ADRF is designed to promote collaboration, facilitate documentation, and provide information to data providers about how data are used.² It currently houses administrative data from 12 state agencies and is under contract with OPRE through the TANF Data Innovation project to provide access to federal TANF and National Directory of New Hires data. Chapin Hall is a partner with NYU on the ADRF. The ADRF has benefitted immeasurably from the experiences of the FSSDC, particularly in recognizing the importance of having sufficient financial resources to address the governance and technical challenges and take advantage of economies of scale.

We did have some success linking and analyzing data. For example, late in the course of our project, we partnered with the HHS Office of Family Assistance (OFA) to compile and organize federally reported TANF microdata for the years 2009 to 2017. We conducted exploratory analyses of the files, explored connections over time, validated cross-sectional data points against the longitudinal history, and completed some basic caseload dynamics analyses. We ultimately developed a longitudinal data file and a report detailing our findings with regard to the data contents, quality, and fitness for various analytic purposes (Wiegand, Goerge, Han, & De La Cruz, 2018). Similar to the process for state agencies, executing the federal DSA for access to these microdata took a

¹ The Federal Risk and Authorization Management Program (FedRAMP) is a standardized approach to data security management for cloud services.

² The ADRF is part of the Coleridge Initiative (<https://coleridgeinitiative.org/>).

significant amount of time. However, creating a repository of these federal data was a much more straightforward process. Partially, this was because we did not seek initially to negotiate permissions and policies to reshare the data with researchers, nor were we working across agency boundaries. We dedicated our own time and resources to understand, document, and organize the data. The data provider was the primary interested party and (as our funder) knew that there would be sustainable resources with which to build on our initial efforts. Even though we only had information about a single program, FSSDC analyses revealed many basic, but important, insights into TANF caseloads that were not well understood without the data resources in place.

Many federally funded programs for which linked administrative data would be most useful (e.g., TANF, Medicaid, SNAP, child support) require state agencies to send administrative microdata to federal agencies. These different data are potentially rich sources for analysis. However, they are often collected for the purposes of measuring federal compliance and may not contain the detail necessary to conduct more complex analyses. In addition, these data emerge from distinct administrative silos, which create data governance problems that complicate efforts to link or connect different types of program data.

Conclusions and Recommendations

Ultimately, ownership and buy-in from data providers is a key driver of the success or failure of any effort to create a repository. Although researchers and funders will naturally benefit from processes to streamline the utility of data for analytic purposes, data providers need to see clear benefits and long-term sustainable operations that justify their time negotiating agreements, drafting governance policies, and preparing data, as well as the inherent risk in allowing a third party to store sensitive data. Potential benefits could include:

- Clear messaging about the purpose and operation of a data repository: The importance of targeted talking points for each stakeholder group and a clear explanation of what that stakeholder can expect to get out of the project, cannot be understated. This includes information that clarifies limitations on live or real-time analyses, as well as emphasizing larger and longitudinal analytic questions.
- The opportunity to streamline existing, burdensome data sharing and governance activities: Systems that reduce demands on staff and agency resources may be attractive to state agencies.
- Assistance with creating and maintaining data documentation: Putting data in a repository requires providing metadata and documentation that currently do not exist in a public-facing format. Generating such documentation may pose too significant a barrier to moving forward. If participation in a repository represents an opportunity for state staff to receive assistance and additional resources (either funding or staff capacity) in creating and maintaining documentation, state agencies may find more incentive to participate.
- Increased data access: Just as researchers see the value in a repository that facilitates easier access to the data needed for research, agencies often have datasets that they themselves want to access: These may be data from other programs on their participants' barriers or other points of context, data on participants' outcomes after they leave the program, or data on individuals who move or take jobs in another jurisdiction (e.g., out of state). If a repository can help facilitate this access, an agency may be much more interested in sharing its own data in turn.

TECHNICAL ASSISTANCE ACTIVITIES TO IMPROVE AGENCIES' USE OF DATA

Our Experiences

Traditional TA activities—such as training and consultation—were the most common activity we conducted across the six years of the FSSDC initiative. We experimented with various forms of TA, ranging from one-on-one assistance to workshops to small groups. The evidence base of essential features and high-quality effective TA practices is still in development (Fixsen, Naoom, Blase, & Friedman, 2005; Katz & Wandersman, 2016), and there is little published literature on TA best practices specific to improving data use in human service agencies. However, resources from other fields, such as health and prevention science, can provide useful insights to guide TA planning and implementation. We aimed to design TA approaches to be responsive to the current competencies of our agency partners and their organizational contexts (Blase, 2009; Katz & Wandersman, 2016). Throughout our experiences and experiments with TA provision, we generally followed a four-phase process: assessment of needs, cooperative planning, delivery of TA, and evaluation (Nemec et al., 1991). In some instances, the process was iterative, moving among phases over the course of the engagement.

Our earliest TA efforts were conducted one-on-one with a handful of pilot states. We used a number of avenues to identify pilot partners: the needs assessment interviews; networking efforts at biannual Research and Evaluation Conference on Self-Sufficiency (RECS) meetings convened by OPRE and at annual meetings of the National Association for Welfare Research and Statistics (NAWRS); and referrals from the American Public Human Services Association (APHSA) and National Association of State Chief Information Officers (NASCIO). Given the capacity of the Data Center and state interest, a set of four initial pilot states were invited to pursue administrative data projects that would receive technical assistance from the FSSDC. Pilots received technical assistance from staff across the FSSDC team but did not receive financial support.

Partnerships with these pilot states extended over the course of four years, though they ebbed and flowed at different times. Ultimately, some pilots were more intensive than others. For all four states, FSSDC provided help understanding the structure, contents, and quality of available administrative data, including restructuring data to be more useful. In addition, FSSDC helped all four states translate policy questions such that they could be answered using the available data. For two states, FSSDC offered advice on internal state data initiatives. Two states also received advice on using FSSDC tools and resources. For one state, the FSSDC facilitated collaborations with researchers from the Family Self-Sufficiency Research Consortium.

In our pilot activities we found that, while we were able to inform ongoing conversations and assist with administrative data challenges, we had relatively few opportunities to use or develop generalized resources, which limited our ability to transfer knowledge to the broader field and to create resources that would persist through staff turnover in the agencies. We attempted to address this limitation by exploring small group technical assistance, with the idea that states working together in small groups could identify issues that were not unique to any single state and might have general solutions. In our pilot project for this form of TA, we partnered with five states to examine the self-sufficiency trajectories of TANF cases, particularly around recidivism and long-term earnings. As part of facilitating that project, we designed a format for the analytic file, shared guidance on that format, and worked with states to answer questions and document nuances of definition. We partnered with states on developing the data, including implementing a generalized R spells

script. Together with the states, we designed two analyses and documented the necessary steps and potential results in detail to share with the states. In two of the five states, we negotiated data sharing agreements and executed the analyses for the states; in the other three states, we supported their internal teams in executing the analyses, including providing support around definitions and analytic decisions, state-specific circumstances, internal data access and governance challenges, and software solutions.

After the small group analyses were completed, we compiled an interview protocol and conducted retrospective interviews with all partners. In the interviews, we sought to understand states' motivations and context for participation, to solicit their feedback on the project and ideas for future similar TA activities, and to capture their experiences using analyses to create impact. For more detail about this project, including the results from the interviews, see the inset case study (p.8-10).

In addition to our customized TA options, we also explored TA offerings for a wider constituency. We hosted two workshops with states and other key stakeholders interested in family self-sufficiency data use. Altogether these workshops engaged a total of 11 states, as well as members of the Family Self-Sufficiency Scholars program and the FSSRC, observers from APHSA and the Pew Charitable Trusts, and the FSSDC team. Workshop sessions focused on administrative data use cases on family self-sufficiency topics, data security, data sharing agreements, managing data, FSSDC tools and resources, formulating questions, and translating analytics into program or policy change.

Participants provided favorable feedback on both workshops. In particular, they appreciated the chance to reserve time in their day-to-day schedules to discuss larger issues and to meet and interact with peers in similar roles at other states. Particularly following the second workshop, participants expressed a strong interest in communicating with other states. Participants wanted to devote more time to state-specific use cases and interactive discussion with other states. They also wanted to create state-to-state peer networks to facilitate future collaboration.

In response to this feedback, the FSSDC developed a LinkedIn group and an occasional email newsletter to connect and inform interested stakeholders. We also sought to schedule follow-up conversations and peer working groups. However, once individuals returned to their offices, their proactive engagement was minimal, and none of these peer initiatives took off. (We also invested relatively lightly in communications or in attempts to build a network.)

Conclusions and Recommendations

Repeatedly during our TA experiences we heard about challenges in maintaining partnerships and meaningful engagements over time. We also heard about challenges getting TA to have an impact, both on outputs, like the ability to generate a report or complete an analysis, and on outcomes, like program effectiveness.

Relationships between TA providers and recipients built on collaboration and trust are an essential component to success alongside the capacity and motivation of TA recipients and their organization to implement innovations (Blase, 2009; Katz & Wandersman, 2016). In our experiences, threats to the persistent engagement of states include staff turnover, limited internal resources, and competing commitments. In states where we engaged with most success, the state partner typically was able to dedicate one or two key individuals to direct the work. It was the case, however, that when those individuals took new roles or otherwise moved away from the work, engagement waned and needed a restart. If we had engagement from senior leadership but not from analysts, there

were often disconnects between plans and execution. Even in the best staffing circumstances, reality remained: government agencies were constantly slammed by new, urgent asks and rarely had time for proactive projects like working to improve data use.

Public agency interest in a new project is often piqued when there is a political or policy development that spurs efforts to build and extend analytic capacity, such as the launch or passage of a new policy or program. At times, it may be fiscal incentives or threats of fiscal penalty that motivate agencies to develop and deploy administrative data capacity. At other times, a case that gets media attention and needs to be addressed in some fashion may spur action.

These changes or conversations create windows of opportunities for successful data partnerships. One challenge in providing TA to states in these circumstances is in having staff and resources ready to capitalize when a window opens. The flexible, extended nature of the FSSDC funding allowed us some range in this project to react and capitalize on opportunities, but funding challenges are not a piece that should be overlooked. Many research firms, academic centers, and think tanks do not have ready funds or capacity to act quickly with state agency partners. External partners often require substantial startup funding and cost overhead recovery in advance of committing staff or time. Often there is a presumption that after a short launch period, new partnerships should be sustainable. Yet, it often is easier to fund new partnerships than to sustain existing partnerships.

Alternately, states may seize on an opportunity when it is presented, as in the context of our small group research project. The presence of the FSSDC, a time-limited external partner, as well as the participation of other states, presented states with a limited opportunity and likely encouraged some to act when they might not otherwise have been primed to partner with an external researcher.

Regardless of a state's immediate motivation for investing in capacity-building activities, all too often the capacity of state agencies to partner and use administrative data lags behind the interest in or mandate to use administrative data. Government agencies need certain data systems and staff capacity in place to move forward on many analytical goals. If there is a significant gap between capacity and aspirations, it potentially takes significant investments for an agency to get to a win, taxing the organization's motivation in the meantime.

Partnerships that persist can be facilitated by the provision of resources or other short-term incentives to help agencies stay invested in the long term. Simultaneously, it is important to start early on managing expectations with regard to the partnership, the demands on the agency, and the feasibility of various goals. TA partners can help agencies recognize the difference between feasible and aspirational questions and help chart a path to how more aspirational questions may be answered.

Even when both sides are able to invest resources in a project and move toward common goals, moving from improving technical outcomes (like the ability to run an analysis) to using data to make an impact on the program and the well-being of participants remains a significant charge. Too often, analytic projects are undertaken to answer questions of interest without consideration as to how the resulting data insights can be applied. It is particularly hard to provide TA for interpreting data, however, because the interpretation and application of analytic results are policy decisions. These decisions reflect not only analytic truths but also the principles and priorities of state leadership. It is not uncommon for discussion and activity around administrative data use to frame such work as objective, unbiased, and removed from

(continued on page 10)

MULTISTATE TANF PROJECT ON SELF-SUFFICIENCY OUTCOMES: A CASE STUDY ON SMALL GROUP TA TO FACILITATE ANALYSIS OF STATE TANF DATA

Emily R. Wiegand and Leah Gjertson

This case study summarizes state experiences and feedback from the TANF Project on Self-Sufficiency Outcomes. This multistate collaborative research project was one of our final FSSDC initiatives, running for just over two years (fall 2016 through the end of 2018). The goal was to measure the employment outcomes of a cohort of TANF leavers. We undertook the project to pilot TA as a small group exercise, combining the personal interaction of one-on-one TA with generalized supports like dataset descriptions, analytic plans, and code samples, tied together with regular peer interactions and collaborations. In this project, we worked jointly with six states (one stopped participating midway through the project).

PROJECT BACKGROUND

From the project's inception, we set out to facilitate a shared research exercise while remaining as flexible as possible. We purposefully did not seek to dictate the terms of the multisite research project or require states to participate in a consistent way. Instead, we worked with the states to jointly refine a set of research questions,¹ select a cohort, define key variables, and ultimately craft a step-by-step analysis plan. In all areas, we sought to "lead from behind," prioritizing the interests and experiences of the states, while also keeping the conversation organized, ensuring consistency across states, and keeping the work moving forward. We experimented with different approaches to soliciting opinions. Often we put forth sample ideas or frameworks for consideration. We conducted three all-state calls to discuss key aspects and (due in part to the challenge of scheduling five states) often scheduled one-on-one calls with states to solicit their feedback and address any state-specific questions or challenges. In one case we had a call with two states to explore a secondary analysis of interest to those states. At one point, when we were looking to narrow our research questions, we fielded a short online survey to solicit information from the states about key outcomes and subpopulations of interest. All TA interactions during the project period occurred remotely via phone conference, often using screen sharing.

The extent of support that we provided to the states varied widely. From the onset, we required that participating states either reserve internal resources to run simple descriptive analyses or execute a data sharing agreement to allow us to conduct the analyses directly. Three of the five states that completed the project did the analysis themselves; in two states, our team did the analysis. In addition to those analyses, we consulted with each of the states about policy and definition questions. In several states, we helped our partners tackle data sharing and permissions processes with other state agencies. We also developed an R script that converts monthly case records (a common data format for state agency records) into spells (a data format well-suited for answering longitudinal research questions). We prepared training and documentation materials for that script and provided technical support in implementation.

States were not required to publish their results or even share results with one another. However, since we heard an interest from all states in benchmarking, we informally shared and compared results during the final group call. We made no attempt to audit analyses to ensure the ultimate accuracy or comparability of those results.

Participating states ranged from very small to very large and included both large rural areas and major metropolitan centers. Both state- and county-administered programs were represented, as were a range of approaches to TANF policy and data analysis. We made no attempt to limit our group to states with similar characteristics.

INTERVIEW METHODOLOGY

After the project was completed, we conducted semi-structured interviews with our contacts in each of the five states that ultimately participated.² Overall, we interviewed ten people. Two were TANF agency leadership who did not have primary research or data responsibilities, four were managers within research or data teams, and four were analysts.³

We asked respondents about state characteristics and analytic priorities, including motivations for participating in our project, feedback on the project and TA provided, feedback on the peer collaboration component of the project in particular, and overall thoughts about using data and analyses to impact policy, programs, and recipients. Below, we summarize key themes that emerged from these interviews.

FINDINGS

State Characteristics and Analytic Priorities

While participating states had a baseline level of data access and expertise in order to participate in the project, respondents discussed both the strengths and weaknesses of their states in using data. With regard to strengths, three states mentioned having accessible data warehouses to facilitate analytic work (although some of the same states cited challenges fully accessing and utilizing these data sources for analytic purposes). Two states talked about the strength of their staff and institutional knowledge with regard to understanding the policy and program context. On the side of weaknesses, three states mentioned challenges in internal coordination and communication, such as standardizing the definitions of key measures and concepts, collaborating across different entities (agencies, jurisdictions, etc.), conducting research and analysis, and communication between analysts and IT. Two states were dealing with challenges matching data between recent data system transitions. Finally, two states cited challenges seeing impacts on programs or policy coming from their work. One respondent mentioned that analytic work felt "piecemeal" rather than holistic; another said, "I feel like [analysts] often don't get utilized enough."

When asked about their motivations for participating in this project, three states emphasized that the topic (employment and self-sufficiency) was important to them and an important thing

¹ Participating states decided to focus on two research questions. The first was a primary analysis of "How does the likelihood of TANF case reopening vary by earnings at case closures?" The second, a supplemental analysis, was conducted by a subset of states. It asked, "To what extent do close reasons related to earnings correlate with independent wage data?"

² A copy of the interview protocol is available in Appendix B.

³ In four of the five states, there were two or more respondents on the phone, so we generally summarize the responses based on the number of states who discussed a topic rather than the number of respondents.

to understand about their programs. One respondent described the decision to participate in a study on these questions as “kind of a no-brainer.” Most of the states had done or considered similar analyses in the past. Two states focused on the opportunity to compare insights and results across states as a unique facet of this project, and one state observed that the project had value in that it provided an external driver to make space for this analysis although “we’re often really buried in the work.”

With regard to other areas of analytic interest, most respondents articulated a desire to go beyond the initial analyses of this project to better understand, as one respondent put it, “how and why” case reopening and employment outcomes occur. They wanted to segment employment and wage outcomes in subgroups based on key characteristics, like geography, educational attainment, and household composition, to understand the characteristics of leavers with different outcomes. Two states also referenced a desire to analyze the combinations of services households need or use, integrated across programs. Finally, two states expressed an interest in looking at employment outcomes using Workforce Innovation and Opportunity Act (WIOA) measures, a current topic of policy discussion.

Feedback on the Project

Overall, states gave very positive feedback on the TA approach. States liked the opportunity to get fresh eyes on their data, questions, and problems. They appreciated the FSSDC’s perspective and subject matter expertise, as well as our “collaborative, conversational approach” and willingness to be flexible as states identified limitations or encountered problems. One respondent noted that the FSSDC team was open from the start about what the project would provide and what the TA parameters were; by contrast, the respondent commented that some data TA efforts overpromise, using “language that doesn’t really match what is truly being offered.”

In particular, states felt the documentation and code prepared during this project were clear and appropriate, resulting in relative ease of implementation. Three states mentioned that the analysis plan documents were straightforward and helpful, and four cited the usefulness of having clear documentation of data needs and definitions available from almost the beginning of the project. Two states spoke positively about the R code (“It was very easy for me to follow those instructions and run the code”) and mentioned that it was being reapplied to expand other areas of work.

The availability of FSSDC staff for near real-time troubleshooting and consultation was also appreciated. Two states mentioned the value of this support. One respondent commented on the willingness of the technical expert to “actually listen to what you are trying to explain rather than trying to continually provide a solution that doesn’t really fit the problem.”

However, states noted that they struggled to make the project a priority. Two states noted the challenge of finding time to work on this project when faced with competing commitments. One suggested that stricter project timelines would have helped them to focus better; the other noted that having more time to complete the analysis would have been helpful.

Other barriers to executing the analysis included data access and approval of the R software. Two states mentioned that access to data ready for analysis, as well as access to comprehensive documentation, can be a barrier to working with external researchers or partners on

this kind of project. (In the states that participated, challenges for this project were fairly limited.) One state experienced some delays getting approval to use R through state IT.

Peer Collaboration

States prized the opportunity to collaborate with peers. At least three states specifically mentioned the value of hearing about policies, processes, and programs in other states, and two specifically mentioned the utility of identifying shared challenges in data and definitions. Three states particularly appreciated the opportunity to share results and benchmark off other states. One state mentioned the usefulness of having a network of peers and a corresponding contact list available after the project.

In fact, states wanted more peer collaboration than we provided in this project. Three states said the project provided only limited contact with other states, and two said interactions with FSSDC staff members impacted their thinking and practice more than interactions with other states. Respondents recommended a variety of potential configurations for a future project, such as:

- More one-on-one contact between states or meetings among smaller groups of states that share specific policies in common, allowing for more nuanced discussion;
- Specifically earmarked peer-to-peer time (i.e., opportunities for discussion between analysts on implementation details, rather than all meetings including managers);
- Allowing more space for introductions and for the group to get to know one another at the beginning of the project, to develop a sense of a network at an earlier stage; and
- Opportunities for face-to-face interaction.

We also asked how large a group of states they thought would work best for a project like this, and whether that group should be restricted to states that look the same on one or more dimensions (e.g., only small states, only states with large urban areas, etc.). Almost all respondents said a group of between five and eight states seemed like the right number. Although three states asserted the value of hearing from states with different perspectives (as one said, “[we] don’t want just states like us”), most also agreed that it would be helpful to work closely with states that share traits in common with their experience. However, states think about similarity across an array of dimensions. These include population (three states), urban/rural distribution (two states), eligibility policy (three states), state- or county-administered status (one state), and diversity/size of immigrant population (one state).

Making an Impact with Data

We asked respondents about the implications and impact this project and other uses of data have for their programs, policies, and participants. In their answers, states identified a number of ways data make an impact. We organize these answers generally according to the type of stakeholder informed by given analyses.

Respondents described using data to inform policymakers, particularly to educate them about common misconceptions about TANF recipients. Examples respondents gave included the perception that the TANF caseload is larger than it is, that TANF recipients are not active in programs and work activities, and that the “problem” is finding jobs (where data suggest the problem may be jobs that do not pay enough). Data also provide TANF leadership with supporting evidence to back up anecdotal

(continued from page 9)

knowledge about differences across subpopulations (such as by household composition or geography) that impact service utilization and results. Overall, respondents from several states commented on the importance of using data to demonstrate to policymakers the realities of poverty. Policymakers are often eager to measure effectiveness or return on investment against difficult concepts like family self-sufficiency. “Sometimes there are risks to performance data without context,” one respondent noted.

Data are also used to support program leaders. For example, targeted analyses can guide policy changes or the design of pilot projects; one state gave the example of designing efforts to improve child care access for TANF recipients. Data can also track operational changes and challenges, like sudden increases in denials or sanctions that may indicate confusion or problems in a local office.

Some data may be useful to frontline staff as well. Examples include using data to develop recommendations to frontline staff about the services or referrals a family may need. These recommendations could improve how the family is served and facilitate faster case processing. Another respondent noted that providing caseworkers with real-time data on work participation enables those workers to more effectively monitor program participants.

Finally, analyzing data helps analysts and researchers learn about data system problems and challenges. Examples raised by respondents include inconsistent coding, how policies like case closings or sanctions are defined and appear in the data, and the accessibility of longitudinal data for analytic purposes. Two respondents noted that this project’s analysis process is helping to inform the design of new data systems or data management infrastructure to better facilitate use of data in the future. Another respondent noted that simply looking at the data in a different format than usual encouraged new insights and taught them new things about their data and programs.

KEY TAKEAWAYS

State feedback from the TANF Project on Self-Sufficiency Outcomes reinforces the lessons we learned in other places about how agencies use data. In particular, states were enthusiastic

about opportunities to do analyses, explore employment outcomes, and collaborate with other states. Respondents routinely see ways where data makes a difference or could make a difference in their state’s practice and policy, and in the lives of the people served. However, availability of data access and staff time are barriers to conducting many analyses. They are also barriers to more deeply engaging with TA opportunities and other states. Analysts struggled to make this project a priority given other urgent agency needs.

A few characteristics of this project helped states to overcome these barriers. Several respondents valued the upfront clarity of the project’s scope, including about the TA supports that would be provided and the data that would be needed. They also appreciated the flexibility and variety of TA approaches. Our goal was always to meet the states where they were; state feedback suggests we succeeded in that.

The project timeline extended over two years from initial solicitation for participation through joint review of the results. We employed a flexible timeframe that allowed time for each state to work through their own challenges and delays to complete the next step of the process, therefore enabling the states to move through the project together. In practice, this meant that, often, months passed between interactions with the full cohort of states (we typically had one-on-one contacts with the participating states during these periods). This flexibility enabled the group to stay on a common timeline. However, it may have also made it difficult to maintain momentum.

States wanted more opportunities to collaborate and interact with peers, including on an in-person basis. It seems that many states would have liked to be pushed to make more time for those interactions, even if that process slowed down the overall project further. (In-person TA would also increase the costs associated with the project.) States would also have valued opportunities for interaction in smaller interest groups of peers who shared key characteristics in common, ideally without sacrificing chances to hear experiences from a larger and more diverse group.

(continued from page 7)

the politics of policymaking. The reality is that where analysis truly makes an impact is in an inherently political space.

While a TA partner must walk carefully around applying a policy interpretation to analytic results, there are critical skills for analytic interpretation for which TA can be provided, such as an understanding of the limitations and biases inherent in administrative data systems. Key considerations that all partners in a data use initiative should discuss include:

- How administrative data work inherently focuses attention and priorities on what is regularly measured or captured in data systems;
- How administrative data do not give voice or insight to parties outside the data collection interaction; and
- Ways that analytic work with administrative data informs decisions about when and how to act, as well as decisions about how to allocate or distribute resources.

Finally, beyond targeted partnerships, states regularly express strong interest in working in groups and sharing knowledge with their peers. It is difficult to get these partnerships beyond the superficial, because 1) again, time and resources are a struggle and 2) there are many state-specific contexts. Due to these many

contexts, quick conversations or examples are more likely to highlight differences (i.e., “why that won’t work for me”) than role models. However, state participants in the multistate TANF project asserted the value of this piece even at the expense of more time or resources and even for diverse groups of states (see the inset case study for more detail).

DEVELOPING SOFTWARE, TOOLS, AND RESOURCES TO SUPPORT DATA USE

Our Experiences

We have approached the idea of providing tools to facilitate data use in many ways over the six years of the initiative, including through tool-building, code sharing, and other technical resources.

As part of the original grant submission to OPRE, the FSSDC proposed to build out an online data tool that administrative data users could access to manage, analyze, and visualize data. Working with NORC’s Data Enclave and Orlin Research, the FSSDC team formed the technical infrastructure to create a secure and intuitive web-based data tool. We developed the data tool to give data users easy access to their administrative data resources. Its functionality allowed users to prepare new measures (e.g., program spell lengths or transitions), create reports, and complete more complex

analyses with ease and without any new technology or software investments. In contrast to similar graphical user interface-based reporting and analytic tools, the tool also had a built-in interface to track and organize metadata about component datasets.

The FSSDC engaged three partner state agencies in pilot projects directly using the tool. We also referenced the tool in conversations about data management systems with several other state partners. Ultimately, despite interest in early demos, we never got buy-in for a comprehensive all-in-one solution—the original vision for the data tool. States questioned the need to tackle the learning curve of a new tool, particularly if they were happy with their current technical capacity for reporting. One state, for example, had had very positive experiences using Tableau. The state saw no reason to learn a new tool that could duplicate only some of that software's functionality. There was no way we could develop software tools within our limited budget that would match the functionality, interface, or overall utility of popular data tools developed in the private market.

Ultimately, we started to question whether it was necessary to continue to build and maintain our own tool. We decided to target the same resources more narrowly at the domain-specific challenges that general data software solutions would never address. Our goals shifted from focusing on the development of a data tool to developing resources that would support data users in a wide range of contexts and data infrastructure. An essential element of this work was recognizing we needed to create modular solutions that addressed specific problems commonly experienced by data users. While not every data user would need all the modular solutions, there were economies of scale to producing resources and tools that would be accessible to a number of data users and be relevant to whichever data platform a user might access. We also realized it was much easier to convince potential users to allocate smaller amounts of time to learning a new tool with a modest learning curve and tangible immediate benefits than a larger, complex tool that did not specifically address an immediate need.

Our first modular resource was a TANF data model developed to address this need. As we worked with agencies and the data tool, we discovered that the format of the data input into the tool tremendously influenced the usability of the result. Certain data formats were more intuitive to analysts and could be input to analyses and visualizations more easily, without significant data transformation.

We ultimately published a brief detailing the data model and the process of putting TANF data (or other monthly benefit data) into this structure for analyses (Wiegand, Goerge, and Gjertson, 2017). We also presented on this model at several conferences. By this point, we had learned that, in most state data systems, data are stored in formats that made sense for performance efficiency and supported data collection and review of single records for case management purposes. Data analyses often require different formats, however, and neither analysts nor database administrators are trained in designing and creating files for analysis.

The TANF Data Model helps states and counties that collect rich data and have analysts or eager research partners on staff but find their data unwieldy and difficult to use. These agencies often have questions that they wish to use the data to answer, but they struggle to define exactly what data are needed to answer those questions and how the data need to be structured. In particular, states often ask questions about program participation spells,

churn, transitions, or longer term outcomes that current data structures do not easily address. This challenge often falls into a gulf between IT staff or consultants and program staff; program staff think in terms of research questions, but these are not easily translated into the technical specifications IT staff need to pull and reformat data.

One key component of the TANF Data Model was the creation of fields to indicate the first month and last month of a continuous period of benefit receipt, or a service delivery spell. Of all the data manipulations done to put data in the data model format, spells creation was the least intuitive. Particularly in a programming language or tool with limited versatility, it can be difficult to create spells, and even more difficult to do so in an efficient way (i.e., in a way that runs in fewer than several hours, even for a moderately sized TANF caseload).

We recognized that creating spells was likely to be a pain point for agencies looking to put their data into our data model. There is no value to states in developing proprietary code to generate spells; it is a purely programmatic process. However, it is specific enough to the domain of public benefits that there are few existing resources. In other words, it was an excellent candidate for our next modular tool effort. We developed scripts in both R and Python to facilitate the creation of data in a spells format from data in a point-in-time format (for example, taking monthly records on benefit receipt and converting them to records for each spells of continuous participation, identified by start and end date).

Our decision to create scripts in R and Python was driven, in part, by our finding that using these languages made it much simpler to write clear and efficient spell code, compared to something like SAS or Stata. Also, as open source languages, R and Python are freely downloadable.³ Of course, in order for government employees to use these tools, their internal information technology offices would need to sign off on the installation. This area of challenge changed significantly over the six years of our project. At its outset, the idea of any open source programming language was anathema to government users. By the final year of our project (2018-19), it was fairly common for state agencies to have had some conversation about R and not that difficult to get it installed. Python was still not very common and a higher hurdle but given the trend of the past few years and the ongoing popularity of data sciences topics, we expect barriers to adoption of both R and Python will continue to decline. Eventually, both should be relatively accessible, even in government.

In order to share these scripts and create a platform for others to share similar resources, we created a repository on GitHub. Although GitHub is little known at present in the social sciences, it is the go-to source for code sharing in open source software, open science, and data science. The repository was a useful platform for sharing code. Furthermore, in introducing the repository to new users, we sought to encourage a new way of thinking in the social sciences—one in which code is a sharable, collaborative product. We incorporated resources on our GitHub repository to introduce users to GitHub and Git (the version control tool that underlies GitHub), as well as to R and Python. After receiving feedback that the raw GitHub repository view was intimidating for new users, we created a GitHub Pages version of the repository, so that at first glance, the repository looked like any other webpage.⁴

In the final years of the project, we used GitHub to publish a resource

³ We also considered SQL but the operations needed for spelling often require database-specific SQL functions, creating generalizability challenges similar to working in the proprietary programming languages.

⁴ To see the GitHub pages version, see <https://chapinhall.github.io/FSSDC>. To see the raw GitHub repository, see <https://www.github.com/chapinhall/FSSDC>.

for administrative data users seeking to understand questions of data quality: the Data File Orientation Toolkit. The toolkit enables data quality assessment of state and local administrative data files. It is designed for an audience of researchers who may be comfortable with programming and datasets but who are new to an administrative data source. These researchers might be new to the unique quality challenges of working with administrative data in general. The Data File Orientation Toolkit helps users quickly become familiar with a data file and navigate research-relevant characteristics common to administrative data. It produces a report assessing an administrative data file across important data quality dimensions, including accuracy, completeness, and comparability among subgroups and over time.⁵ The analyses in the toolkit are informed by best practices from the literature, focusing on exploratory data analysis and data visualization to detect notable patterns in the data file. The Data File Orientation Toolkit is based in R Markdown, but is designed for use by individuals who may not be familiar with R.

Conclusions and Recommendations

Our work developing tools and resources for data users revealed several insights that are relevant to future work to improve computing and analytic environments for administrative data users. When developing tools and resources, we found it challenging to successfully encourage adoption and sustained interaction with new tools. Agency staff often face urgent deadlines and competing priorities; they have minimal time and attention for new things, even when they think the new thing is important and might significantly benefit them down the road. Analysts may also assume that resources that use new programming languages, or do not come with a graphical user interface, are not meant for them because they do not have computer science training and do not think of themselves as programmers. Finally, even if the potential user is excited and motivated, tools that require the installation of new programs (or even just updated browsers) frequently require IT installation and support that may create significant delays and undermine momentum.

To promote adoption of these new tools, therefore, we sought to build off existing systems or arrangements that were familiar to the agency and data user wherever possible. For example, we could build faster, more efficient scripts in Python than in R. Yet, because R is an analytic language more familiar to our partners and their IT staff, we chose to create more scripts in R than Python. Furthermore, RStudio provides a graphical user interface that reduces the learning curve for analysts used to working with standalone software rather than command line programming languages. As a result, we used R, even if it meant the script ran for several hours longer than it would have in Python.

We also sought to lower the learning curve for any new tool or resource. Simpler tools or resources that address a single challenge, rather than “one size fits all” solutions, are of benefit here. They require learning fewer features. Simple documentation and multiple modes for learning new tool functionality (e.g., written resources, videos, and live trainings) also help lower learning curves. In our work, we presented resources at professional conferences and meetings; on these occasions, the users we wanted to reach had already set aside time and attention to learn new things.

We learned that the premise “if you build it, they will come,” rarely applies to the world of administrative data. We found that even the best tools or supports need to be aggressively marketed and demonstrated and must be accompanied by excellent documentation and training materials before they see widespread adoption among state agencies.

Role models or champions of successful tool adoption can be instrumental in fostering broader awareness and adoption. We often used the experience of an early adopter to demonstrate a tool's relevance to a particular context. Testimonials and endorsements from early adopters also demonstrate that peers are able to use a given tool. Peer effects are particularly powerful for overcoming hesitation around learning new programming language or adopting tools without a graphical user interface. Successful implementation provides success stories and proof of concept for data users to share with senior leadership and IT staff. We particularly saw this experience in the context of R. Some analysts initially assumed their IT would not approve and install an open source software package. When we told them that other states had, the analysts reconsidered. In some cases, they connected with these other state users to learn more about how the successful agency had demonstrated their need and addressed IT concerns.

Of course, all the best adoption practices in the world will not matter if the tool doesn't meet a demonstrated need. Challenges in using data are rarely just an issue of having the right tools or software; in fact, there is a proliferation of non-domain-specific software solutions to data problems. Given the learning curve for any new software, it is extremely important for potential users to understand why the new solution is needed and how they stand to benefit from investing in set-up time. Similarly, the developer of any solution must understand the problem appropriately, including investing time to determine whether a new tool is required or an existing solution can be more easily adapted. Data access and data preparation are complex issues with a lot of highly customized components that can't be easily generalized or shortchanged.

In the end, our success working with partners occurred when we listened to their needs. This means that any software, code, or resource solution should directly address gaps and challenges identified through needs assessment and technical assistance activities. Tools are not a way to skip past the due diligence required to have a deep understanding of computing and analytic needs. Successful tools are well-crafted to fit a specific problem at hand.

CONTINUOUS LEARNING ABOUT NEEDS, BARRIERS, AND CHALLENGES AROUND ADMINISTRATIVE DATA

Our Experiences








Throughout our six year initiative, we intentionally sought opportunities to learn and share lessons learned regarding needs, barriers, and challenges facing state agencies and researchers who use administrative data. Learning efforts began at the start of our initiative with the 2013 needs assessment. To examine the administrative data resources available and the realities of how agencies use administrative data to inform policy or practice, the Data Center team conducted interviews and focus groups with nearly 100 key stakeholders in family self-sufficiency programming and research. These stakeholders included state agency senior staff, public assistance program administrators, external researchers, members of relevant professional associations, and leaders of advocacy groups. Questions focused on current administrative data capacity, challenges, and needs. In addition to informing initial Data Center strategy for developing tools and resources, we published the results of the needs assessment in a report (Weigensberg et al., 2014) and a subsequent journal article (Allard et al., 2018).

As we worked on the analysis and dissemination of the needs assessment findings, our other activities, especially technical assistance, continued. As a result, we had a chance to see and

⁵ We prepared a companion paper summarizing existing research on administrative data quality and outlining this taxonomy in more detail (Seeskin, Ugarte, & Datta, 2019).

FSSDC LESSONS LEARNED APPLIED BY AUDIENCE

Our findings repeatedly emphasize the need to approach this work with...

	 CREATIVITY	 FLEXIBILITY	 LEADERSHIP
ADMINISTRATORS AND AGENCIES 	Start by thinking about questions, not solutions. Answers should create opportunity for multiple solutions.	Learn from failure. Recognize that where you think you need to build capacity or focus improvements may not be what you actually need.	Secure buy-in from agency leadership to support stability amidst staff turnover and competing priorities. Build on small wins.
TA PROVIDERS 	Be creative and adaptable, willing to try multiple approaches and adjust mid-stream. Look for how a problem has been addressed in other places or with other tools.	Meet people, agencies, and technologies where they are at. Ensure partner questions guide activities.	Deploy clear messaging about what TA efforts do and do not provide, about the benefits TA brings for the partner, and about what the partner will need to invest.
FUNDERS 	In funding data use initiatives, incentivize experimentation, and accordingly seek to measure success and failure and capture lessons learned.	Enable flexibility in how TA is designed and let scope change to meet agency needs. Intentionally manage global goals and daily tasks.	Provide prolonged, sustainable funding to support healthy and persistent partnerships. Follow progress on hard challenges and avoid becoming path dependent.
DEVELOPERS 	Start with simpler resources (e.g., minimum viable products) to address single challenges rather than “one size fits all” solutions.	Keep learning curves as low as possible and build off existing areas of familiarity if possible.	Seek role models of successful adoption (peer adopters are most compelling).

understand, firsthand, agencies’ direct experiences: what they told us when first asked, what they told us when we asked the question differently, and what we actually found was needed when we partnered in a hands-on way. We repeatedly learned how difficult it is to understand what is truly needed to increase data use capacity. There were discrepancies between what people said they needed and what they actually needed. There were also discrepancies between our initial assessments of what was needed in a particular agency environment and what was ultimately needed.

We completed a similar body of work, with a more pinpoint focus, in an assessment of capacity, opportunities, and open challenges on record linkage. Record linkage, the process of identifying the same individual in multiple data sources in the absence of a high-quality, shared identifier, is a crucial consideration in family self-sufficiency research. In that work, it is necessary to combine data sources from multiple programs or jurisdictions that are not stored in a common data system. We conducted a series of interviews with experienced practitioners about commonly linked family self-sufficiency data sources and the record linkage process and challenges (Wiegand & Goerge, 2019c). We also reviewed the record linkage methodological literature and conducted a scan of technical and commercial options to conduct linkage (Wiegand & Goerge, 2019b). Ultimately, we identified a series of defining considerations for this record linkage use case (Wiegand & Goerge, 2019a). We produced two reports and a series of recommendations for maintaining standards of quality as the integration of state and local data sources becomes increasingly prevalent.

Conclusions and Recommendations

Throughout the project, we were reminded as to how important it is to operate as a learning organization that is primed to be in a perpetual state of inquiry around data needs and capacity. In addition to being open to new ideas and ways of thinking, we also found the need to help translate initial conceptions of barriers or obstacles into actionable solutions. Sometimes, we never found the

right lever. Yet, by creating interactions and modes of inquiry that constantly asked about data partner needs, we ensured a higher probability of finding multiple and creative solutions.

Throughout our learning and work, the challenges identified early on remained key challenges in our work with states. We did learn, however, that there are many potential ways to reduce barriers to data use. Which remedy will promote exemplary data use in any specific context is not always clear. The presenting challenge to data use may not be immediately identifiable to an agency or particular analyst. It also often is the case that agencies can’t pinpoint where processes lag or where capacity is needed.

The work of providing TA and partnering with data users should reflect these realities. Attempts to document barriers to data use and capacity building needs should be mindful that there is uncertainty. External researchers should continuously ask about challenges, finding different and creative ways to pose questions to different actors at different points in the data use process. Before acting, external partners also should seek detailed evidence and use cases, and then be willing to reconsider initial assumptions. Thinking of work to advance data use in this manner underscores how experimentation, reflection, and flexibility are key ingredients to successful partnerships. There are not “one size fits all” strategies.

OVERALL CONCLUSIONS AND RECOMMENDATIONS

Apart from the detailed conclusions related to distinct areas of work, we found several key themes emerge consistently across the Data Center initiative. See “FSSDC Lessons Learned Applied by Audience” for a summary of these key themes applied to different stakeholders.

First, it is clear that support from leadership is necessary to make and sustain progress around data use. Support from leadership can mean building capacity, supporting analysts, developing an agenda

for analysis, articulating questions, and incorporating analysis into decision making. Since there is no structural requirement for being a data-driven organization, that will only happen through the active support of leadership. Once developed and funded, data functions in agencies can show benefits, but examples of this are rare.

Second, our experience shows that flexibility and experimentation are crucial. The application of data analysis to policy in state and local human service agencies is a topic still being developed; TA for this topic is even newer. It is difficult to identify exactly what supports are needed, and these needs sometimes shift. It is important to build honest sharing of needs and barriers encountered into efforts like the FSSDC, in conference presentations, and in reports like this one.

Our work underscores the variety of complex reasons that make effective data use difficult. Data users should be guided to see that their problems are not unique, are not easy, and are not solvable by an easy technical solution. To the extent that data users can segment the larger problems into small, manageable areas for intervention, there are resources available to help overcome those hurdles and move toward better use of data.

As is common, we found a tension between providing customized or one-on-one TA and more generalized TA or resources. Providing customized resources or one-on-one TA can make the most immediate impact for a partner but may not build capacity that is sustained after a partnership or after a key agency staff person leaves. More generalized TA or resources can be difficult to apply to individual circumstances and may have learning curves that turn off potential users but have the potential to build lasting capacity.

Finally, the technology landscape is shifting. For example, technologies that once created significant barriers, like cloud storage of data or analytics in R or Python, are becoming increasingly familiar and commonplace in state government. Any effort to support or advance the use of administrative data must operate in a continuous state of learning and adaptation to both changing needs and changing possibilities.

APPENDIX A: FAMILY SELF-SUFFICIENCY DATA CENTER PRODUCTS

This appendix lists all publications and other products created by the FSSDC team over the course of the project.

Resources for Working with Administrative Data

FSSDC GitHub Page: <https://chapinhall.github.io/FSSDC/>

Code snippets and software tools, including spell scripts and the Data File Orientation Toolkit, are hosted on the FSSDC's GitHub page.

"Family Self-Sufficiency Data Center: Creating a data model to analyze TANF caseloads" (Wiegand et al., 2017)

This brief provides an overview of the TANF data model and walks through two examples of transforming state data into the data model.

"Constructing a Toolkit to Evaluate the Quality of State and Local Administrative Data" (Seeskin et al., 2019)

This journal article describes the process of developing the Data File Orientation Toolkit, including an in-depth discussion of dimensions of data quality in administrative data sources.

Research on Understanding Needs for Improving Data Use

"Family Self-Sufficiency Data Center: Needs Assessment Report" (Weigensberg et al., 2014)

This report summarizes findings from the FSSDC needs assessment.

"Agencies' Use of Administrative Data for Improved Practice: Needs, Challenges, and Opportunities" (Allard et al., 2018)

This journal article describes the needs, challenges, and opportunities of public agencies seeking to better use data.

Other Publications

"Understanding and Using Federal TANF Characteristics Data for Longitudinal Analysis" (Wiegand et al., 2018)

This report summarizes the FSSDC's analyses and findings of the federal TANF data extracts.

"Using and Linking Administrative Datasets for Family Self-Sufficiency Research" (Wiegand & Goerge, 2019c)

"Record Linkage Innovations for the Human Services" (Wiegand & Goerge, 2019b)

"Recommendations for Ensuring the Quality of Linked Human Services Data Sources" (Wiegand & Goerge, 2019a)

These three reports discuss the characteristics and challenges of linking state and local administrative datasets to address questions of family well-being and the effectiveness of family self-sufficiency programs.

APPENDIX B: MULTISTATE TANF PROJECT ON SELF-SUFFICIENCY OUTCOMES INTERVIEW PROTOCOL

Introductory

Why did you choose to participate in this project? Have you done an analysis like this one before, or were there plans to do a similar analysis before the initial FSSDC contact?

What do you think are your agency or state's strengths and weaknesses in using data?

Process

What parts of the work (data access, data preparation, analysis, defining the question, etc.) were particularly easy or difficult for your state? Did this align with your expectations?

About how long did the project take (both active staff time and "waiting")?

What (if anything) would you approach differently if you knew what you know now at the start of the project?

What was most useful about the documents and scripts the FSSDC team prepared and shared? What about these materials was not useful or did not work well?

Based on this and other experiences you've had, what works well in providing technical assistance around data use? (This includes format, timeline, process, etc.) What is missing?

Working with Other States

If you were attempting to do this analysis on your own (i.e., not with other states), how might you have approached some of the questions or definitions differently?

What have you learned through collaborating with other states about some of the measurement and definition challenges that are the same across states, or that are unique to your state?

Do you have an example of something you learned or thought about differently following a conversation with another state in the group?

If you were doing something like this again, would you want to do it with more states, fewer, or the same amount? Would you prefer to work with states that were similar to you in some way (policy, data expertise, something else)?

Implications for Policy and Research

What was the most useful part of this exercise for you? What has been the biggest gain for your state to come out this project?

What about the exercise could be changed or adapted to make it more useful?

Did your agency adopt or change any practices, either in policy/practice or research/analysis capabilities, based on this project?

How do you think about connecting findings about earnings and employment outcomes to policy and program administration? In other words, once you have the measurements, how do you use them to improve your programs?

This exercise focused on benefits receipt and earnings outcomes for TANF recipients. What analysis topics are of particular interest to your agency?

What else should we be thinking about for doing collaborations like this again in the future?

REFERENCES

- Allard, S. W., Wiegand, E. R., Schlecht, C., Datta, A. R., Goerge, R. M., & Weigensberg, E. (2018). Agencies' use of administrative data for improved practice: Needs, challenges, and opportunities. *Public Administration Review*, 78(2), 240-250.
- Blase, K. A. (2009). Technical assistance to promote service and system change. *Roadmap to effective intervention practices #4*. Tampa, FL: University of South Florida, Technical Assistance Center on Social Emotional Intervention for Young Children.
- Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59, 1-12.
- Doar, R., & Gibbs, L. (2017). *Unleashing the power of administrative data: A guide for federal, state, and local policymakers*. Washington, DC: Results for America. Retrieved from <https://results4america.org/wp-content/uploads/2017/10/Unleashing-the-Power-of-Administrative-Data.pdf>
- Duran, F., Wilson, S., & Carroll, D. (2005). *Putting administrative data to work: A toolkit for state agencies on advancing data integration and data sharing efforts to support sound policy and program development*. Farmington, CT: Child Health and Development Institute of Connecticut. Retrieved from <https://www.researchconnections.org/childcare/resources/9079/pdf>
- Federal Data Strategy Team. (2019a). *Draft 2019-2020 federal data strategy action plan*. Retrieved from <https://strategy.data.gov/assets/docs/draft-2019-2020-federal-data-strategy-action-plan.pdf>
- Federal Data Strategy Team. (2019b). *Practices for leveraging data as a strategic asset*. Retrieved from <https://strategy.data.gov/assets/docs/federal-data-strategy-practices.pdf>
- Fixsen, D. L., Naoom, S. F., Blase, K. A., & Friedman, R. M. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).
- Goerge, R. M. (2018). Barriers to accessing state data and approaches to addressing them. *The Annals of the American Academy of Political and Social Science*, 675(1), 122-137.
- Green, B. L., Ayoub, C., Bartlett, J. D., Furrer, C., Von Ende, A., Chazan-Cohen, R., & Nygren, P. (2015). It's not as simple as it sounds: Problems and solutions in accessing and using administrative child welfare data for evaluating the impact of early childhood interventions. *Children and Youth Services Review*, 57, 40-49.
- Hotz, V. J., Goerge, R., Balzekas, J., & Margolin, F. (1998). *Administrative data for policy-relevant research: Assessment of current utility and recommendations for development*. Report of the Advisory Panel on Research Uses of Administrative Data of the Northwestern University/University of Chicago Joint Center for Poverty Research. Evanston, IL: Joint Center for Poverty Research.
- Katz, J., & Wandersman, A. (2016). Technical assistance to enhance prevention capacity: A research synthesis of the evidence base. *Prevention Science*, 17(4), 417-428.
- Krosnick, J. A., Presser, S., Fealing, K. H., Ruggles, S., & Vannette, D. L. (2015). *The future of survey research: Challenges and opportunities*. The National Science Foundation Advisory Committee for the Social, Behavioral and Economic Sciences Subcommittee on Advancing SBE Survey Research. Retrieved from: https://www.nsf.gov/sbe/AC_Materials/The_Future_of_Survey_Research.pdf
- Lin, V., King, C., Maxwell, K., & Shaw, S. (2018). *Opportunities through state agency research partnerships for using administrative data to support early care and education*. OPRE Research Brief #2018-53. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Retrieved from https://www.acf.hhs.gov/sites/default/files/opre/opportunities_state_agency_research_partnerships_admin_6_18_18_b508.pdf
- Maxwell, K. (2017). *Issues in accessing and using administrative data*. OPRE Report #2017-24. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services (OPRE). Retrieved from https://www.acf.hhs.gov/sites/default/files/opre/issues_in_accessing_and_using_administrative_data_3617_508.pdf
- National Academies of Sciences, Engineering, and Medicine. (2017). *Innovations in federal statistics: Combining data sources while protecting privacy*. Washington, DC: National Academies Press.
- National Forum on Education Statistics. (2012). *Forum guide to supporting data access for researchers: A state education agency perspective*. (NFES 2012-809). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from <https://nces.ed.gov/pubs2012/2012809.pdf>
- Nemec, P. B., Forbess, R., Cohen, M. R., Farkas, M. D., Rogers, E. S., & Anthony, W. (1991). Effectiveness of technical assistance in the development of psychiatric rehabilitation programs. *Journal of Mental Health Administration*, 18, 1-11. doi: 10.1007/BF02521128.
- Office of Planning, Research and Evaluation. (2016). *Using administrative data in social policy research*. OPRE Report #2016-62. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Retrieved from https://www.acf.hhs.gov/sites/default/files/opre/promises_and_challenges_20160809_508_compliant.pdf
- Penner, A. M., & Dodge, K. A. (2019). Using administrative data for social science and policy. *The Russell Sage Foundation Journal of the Social Sciences*, 5(3), 1.
- Schroeder, R. (2014). Big Data and the brave new world of social media research. *Big Data & Society*, 1(2), 1-11.

Seeskin, Z. H., Ugarte, G., & Datta, A. R. (2019). Constructing a toolkit to evaluate the quality of state and local administrative data. *International Journal of Population Data Science*, 4(1), 1-11.

Weigensberg, E., Schlecht, C., Wiegand, E., Farris, S., Hafford, C., Goerge, R., & Allard, S. (2014). *Family Self-Sufficiency Data Center: Needs assessment report*. Chicago, IL: Chapin Hall at the University of Chicago. Retrieved from https://www.chapinhall.org/wp-content/uploads/FSS_Data_Center_Needs_Assessment_Report_Final_0.pdf

Wiegand, E. R., & Goerge, R. M. (2019a). *Recommendations for ensuring the quality of linked human services data sources*. Washington, DC: Family Self-Sufficiency and Stability Research Consortium.

Wiegand, E. R., & Goerge, R. M. (2019b). *Record linkage innovations for the human services*. Washington, DC: Family Self-Sufficiency and Stability Research Consortium.

Wiegand, E. R., & Goerge, R. M. (2019c). *Using and linking administrative datasets for family self-sufficiency research*. Washington, DC: Family Self-Sufficiency and Stability Research Consortium.

Wiegand, E., Goerge, R., & Gjertson, L. (2017). *Family Self-Sufficiency Data Center: Creating a data model to analyze TANF caseloads*. Washington, DC: Family Self-Sufficiency and Stability Research Consortium.

Wiegand, E., Goerge, R., Han, S., & De La Cruz, E. (2018). *Understanding and using federal TANF characteristics data for longitudinal analysis*. Chicago, IL: Chapin Hall at the University of Chicago.