
Family Self-Sufficiency Data Center: Needs Assessment Report

Elizabeth Weigensberg

Colleen Schlecht

Emily Wiegand

Sherri Farris

Carol Hafford

Robert Goerge

Scott Allard

**Report to the U.S. Department of Health and Human Services,
Administration for Children and Families, Office of Policy,
Research and Evaluation Award# 90PD0272**

Family Self Sufficiency Data Center,

Principal Investigators:

Robert Goerge, PI

Scott Allard, Co-PI

Family Self-Sufficiency Data Center: Needs Assessment Report

Elizabeth Weigensberg, Colleen Schlecht, Emily Wiegand, Sherri Farris, and Carol Hafford

Recommended Citation

Weigensberg, E., Schlecht, C., Wiegand, E., Farris, S., Hafford, C., Goerge, R., & Allard, S. (2014). *Family Self-Sufficiency Data Center: Needs Assessment Report*. Chicago: Chapin Hall at the University of Chicago.

© 2014 by Chapin Hall at the University of Chicago

1313 East 60th Street

Chicago, IL 60637

ISSN:1097-3125

Acknowledgments

The Family Self-Sufficiency Data Center needs assessment and report were supported by funding from the U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. Much appreciation is offered to the numerous state and local public agency administrators, researchers, and other professionals who provided input through our needs assessment process. Thanks are also due to Bob Goerge, Scott Allard, Tim Mulcahy, Pat Ruggles, Catherine Ruggles, Matt Borus, and Michelle Derr who provided input and feedback throughout the process.

Table of Contents

Executive Summary	1
Definition and Concept of Family Self-Sufficiency	2
The Use of FSS Data	2
Purpose of FSS Data Use	2
Capacity to Use FSS Data	2
Data Use and Sources	3
Longitudinal and Linked Data.....	4
Data by Source	4
Data Quality.....	5
Research Needs	6
Feedback for the FSS Data Center	7
Better Use of Data	7
The Role of the Data Center.....	7
Technical Assistance, Training, and Peer Engagement.....	7
Potential Implementation Challenges.....	7
Translating Findings into Actionable Steps.....	8
Next Steps.....	8
Introduction.....	9
Family Self-Sufficiency Data Center	9
Needs Assessment Overview.....	10
Topics Included in Family Self-Sufficiency.....	10
Current Use of Data and Research	10
Data and Assistance Needs.....	10
Methods	12
Sample	12
Focus Groups.....	13
Interviews	13
Analysis	13
Concept of Family Self-Sufficiency	14
Defining Family Self-Sufficiency	14

Overview of the Use of Family Self-Sufficiency Data	17
FSS Data Used for Reporting and Performance Management	17
FSS Data Used for Research	19
Capacity and Challenges Using FSS Data.....	19
Agency Staff Capacity to Use Data.....	19
Agency Data Use Capacity.....	22
Data by Type.....	25
Longitudinal and Linked Data.....	25
Longitudinal Data.....	25
Linked Data	26
Data by Source	27
TANF Data.....	27
Food Assistance Data	29
Unemployment Insurance, Earnings, and Employment Data.....	29
Medicaid and Health Data.....	32
Education and Training Data.....	34
Child Care Subsidy and Child Support	35
Child Welfare	36
Criminal and Juvenile Justice Data	37
Earned Income Tax Credit and Tax Data.....	37
Census and National Survey Data	38
Original Data Collection	38
Other Data	40
Data Quality	41
Data Entry and Accuracy Concerns.....	41
Incomplete Data	42
Understanding the Context of Data Collection.....	42
Timeliness.....	43
Addressing Quality Concerns.....	43
Research and Data Needs.....	45
Research Questions	45

Data and Tools Needed to Address Research Questions	47
Data Needed	47
Tools Needed.....	48
Feedback for the FSS Data Center	49
Better Use of Data in FSS	49
Role for the Data Center.....	50
Training and Technical Assistance.....	52
Implementation Challenges	53
Other Aspects	54
Implications and Takeaways for Informing the Data Center	56
Translating Findings into Actionable Steps.....	56
Data Access and Permissions	57
Scope and Types of FSS Data	57
Data Security, Storage and Documentation.....	58
Data Management, Linking, and Analysis	58
Training and Technical Assistance.....	59
Promote Peer Engagement	60
Next Steps and Timeline	60

List of Tables

Table 1. Summary of FSS data use and accessibility among Needs Assessment interviewees.....5

Executive Summary

The Family Self-Sufficiency (FSS) Data Center is funded by the U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. The Data Center's primary goal is to enhance the use of administrative and survey data to answer policy-relevant questions surrounding family self-sufficiency and well-being. The Data Center pursues this goal by working to increase the amount and quality of data available for FSS research. In addition to making existing and new data sources easier to access, the Data Center will work with policymakers and researchers to ensure data can be stored and analyzed in a secure environment. The Data Center will provide an empirical, data-driven research platform for researchers, policymakers, and administrators to use in answering fundamental policy and program questions, and will build knowledge that will ultimately be translated into better policy and practice.

The objective of the needs assessment is to obtain input from key stakeholders—primarily from potential Data Center users—regarding their current capacity, challenges, and needs with FSS data in order to learn how the Data Center can help meet those needs. Specifically, the needs assessment solicited information in three areas: (1) the topics included in FSS, (2) the current use of data and research around FSS, and (3) the data and assistance needed among the FSS data user community.

To achieve the objective of the needs assessment, the team conducted interviews and focus groups with key FSS stakeholders, including program administrators, researchers, and other interested groups, such as professional associations and advocacy groups. Most interviews were conducted via telephone, while the focus groups were conducted in person. Interviews and focus groups took place between February and May 2014. A total of 96 people participated in either an interview or a focus group.

Definition and Concept of Family Self-Sufficiency

Our interview respondents defined FSS in different ways. The needs assessment attempted to capture a complete picture of the diversity around the concept of FSS that is in use among administrators, researchers, and other data users. It also attempted to explore the interviewees' notions of how FSS can be shaped by local and familial context. Interviewees included the following topics and program areas in their definitions of FSS: Temporary Assistance for Needy Families (TANF), the Supplemental Nutrition Assistance Program (SNAP) and other food assistance programs, Unemployment Insurance, child care subsidies, housing supports, job training, the Earned Income Tax Credit (EITC), Medicaid and health services, education, child welfare, criminal justice, domestic violence, substance abuse, economic and community development, transportation, and assets. More broadly, FSS is characterized by income supplements, work supports, and a variety of areas that contextualize families' financial stability and general well-being. Knowledge of these contextual factors allows analysts, case managers, and researchers to better understand what may be shaping observed behaviors and more effectively target programs to maximize impact.

The Use of FSS Data

Purpose of FSS Data Use

State and local agency administrators were asked to describe their agencies' current use of FSS data. They described a wide array of applications representing varying degrees of sophistication in data use. Agencies use data for reporting, performance management, and quality control purposes including federally mandated reporting, internal metrics or scorecards, and fraud prevention and program integrity work. A small number of agencies also use data internally to answer research questions that inform policy planning, generally by helping to understand the impact of a new or proposed policy change. More commonly, state and local administrators interviewed mentioned undertaking partnerships with external researchers or organizations to fully utilize the data that can address their FSS questions and populations. These partnerships focused on evaluating programs, understanding populations, and estimating policy impacts.

Capacity to Use FSS Data

How agencies and researchers use data is determined not only by what they seek to know but, perhaps more importantly, by their ability to access and analyze the necessary data. Interviewees provided dramatically different accounts of state and local agencies' capacity to use data. About one-third of administrators interviewed described having dedicated research or evaluation staff members within their agency. Another one-third of administrators came from agencies with no formal research staff, but where

one or more staff members are dedicated to data management or reporting. A handful of interviewees described organizations that lack even that capacity. In these agencies, program staff and administrators do their own reporting or data analysis, depending on their time, skill, and inclination. Data management and reporting in these agencies often falls on the IT staff.

In their work, interviewees indicated they addressed wide-ranging research and performance management questions. This requires access to datasets from multiple agencies in many areas, and requires that the data be shared and integrated. Some elements present in state and local agencies aid administrators and researchers in access, sharing, and data use, while others serve as additional barriers to these functions. Administrators and researchers described successful collaborations and data sharing arrangements and they repeatedly emphasized the importance of relationships in making these arrangements work to aid data sharing and access. Organizations that work together frequently and perform mutually beneficial work tend to have less trouble negotiating data sharing agreements. Additionally, having clear data sharing policies typically facilitates the sharing of data across departments.

Legal restrictions are a particular challenge in sharing data. Several interviewees mentioned the difficulty in obtaining permissions to share their own data, working through data sharing agreements, and obtaining consent in accessing data from other agencies. State data sharing policies, even for sharing between departments in the same state, are often unclear or, if clear, extremely stringent. Restrictions on sharing data that contains identifiers are even stricter, which increases the difficulty of combining data from multiple sources.

Resource limitations, including limited staff or system capacity, also impact the possibility of data sharing. If the agency lacks the resources or technical ability to extract the necessary data, sharing is not possible. Additionally, even after the data is shared, staff time is needed to answer questions for researchers who lack knowledge of the agency or its data. Another limitation can be that researchers are interested in data that is simply not collected, at least not in administrative sources, like data on untaxed earnings.

Data Use and Sources

The main part of the needs assessment consisted of interviewers engaging administrators and researchers in detailed discussions about the data they use and have access to in their work. Interviewers also asked about the data administrators and researchers want but do not have available or use to the extent they would like.

Longitudinal and Linked Data

Two key sources of data for FSS analyses are longitudinal data and linked data—data from different agencies, programs, or nonagency sources that have been connected at the individual level. About half of interviewees had access to a data warehouse or another source of longitudinal data to draw comparisons and track participants over time. Many of the questions that both agency administrators and researchers hoped to use FSS data to answer require access to longitudinal data and the capacity to work with and analyze such data. About half of interviewees worked with some kind of integrated database—a database, data hub, or data warehouse in their agency that integrates multiple data sources—and others had linked pairs of datasets on an ad hoc basis. Access to more and different linked data was a top priority for almost everyone interviewed, however. Respondents wished to better understand the full circumstances of a family by integrating as much data as possible about the family’s needs and resources.

Data by Source

Table 1 below summarizes interviewees’ descriptions of how frequently specific data types are used and wanted and how difficult they are to acquire.

Table 1. Summary of FSS Data Use and Accessibility among Needs Assessment interviewees

Data Type	Use	Desirability (want, don't have)	Accessibility
TANF	Extremely frequent	Low (most have basic TANF data—but want linked or longitudinal variants)	Easy
Food Assistance/SNAP	Extremely frequent	Somewhat low (a few researchers want, most administrators have)	Moderately easy (a few researchers mention legal hurdles)
Unemployment insurance/earnings/employment	Frequent (with significant limitations)	High (for comprehensive data available at individual level)	Somewhat difficult (most comprehensive sources tend to be tax related and highly restricted; sometimes no one source has all the data needed)
Medicaid and health data	Frequent (Medicaid eligibility)	Moderate	Somewhat difficult (HIPPA and privacy restrictions make challenging to get individual level data)
Education data	Rare	Somewhat high	Difficult
Child care subsidies and child support	Sometimes	Somewhat low	Moderately easy (may be challenges linking to other data)
Child welfare	Sometimes	Somewhat low	Moderately easy
Criminal justice	Sometimes	Somewhat low	Moderately easy
EITC and tax data	Rare	Moderate (high interest when overlapped with earnings data)	Very difficult
Census and national survey data	Sometimes	Low (moderate interest in linking this data)	Very easy (but very difficult to get identifiers for linking)
Original data collection	Frequent	Low (no knowledge of what other original data exists, so no interest in accessing)	Somewhat difficult (new original data is expensive and time-consuming to collect; existing original datasets are often restricted by IRB or budget limitations)

Data Quality

Regardless of the types of data states, localities, and researchers used or emphasized that they wanted to use in their work, there are some cross-cutting concerns regarding data quality that emerged in our needs assessment discussions. Interviewees had particular concerns about data entry errors and inconsistencies, especially in cases where data is entered by a large group of people, such as frontline staff workers who are short on time. Data that are incomplete raise quality concerns for researchers and administrators using

FSS data. Datasets may be missing certain populations; several interviewees mentioned the difficulty presented by program participants who move across state lines, effectively splitting the data on their program history from their eventual outcomes.

Given these data quality experiences in their own programs, most interviewees were very concerned with understanding the context in which the other data they accessed had been collected. In addition to knowing where data quality issues might be expected, this kind of context can help users understand why fields are missing, what data points have been or should be recoded, and what data each field truly captures. Respondents were also concerned with the timeliness of data. Some data sources have time lags in reporting or are only available annually. These restrictions limit the usability of the data for certain analyses.

Research Needs

Throughout conversations during the needs assessment, states, localities, and FSS researchers mentioned an abundance of research questions they would like to be able to address and study. These can be grouped into 10 broad questions:

- How effective is a specific program and what are some impacts and outcomes on clients?
- How does receipt of multiple public benefits impact a family and what is the composition and impact of those multiple programs?
- What are the outcomes of individuals that have left TANF?
- What are the outcomes for special populations engaged in various FSS programs?
- What are the characteristics of the population in a specific program and what facilitates or impedes their success?
- What are the impacts of economic and policy changes?
- What is a living wage and what is included in living self-sufficiently?
- What are the differences among the various groups of participants?
- What employment supports and benefits most commonly lead to retention and achieving self-sufficiency?
- How does a program compare to its equivalent in other states?

Interviewees suggested a variety of data that would be needed to answer these questions. They particularly mentioned longitudinal data that captures long-term outcomes, integrated data across human service systems to capture multiple program interaction, measures of family well-being, and data from other states for benchmarking. In terms of tools, respondents felt a data warehouse containing longitudinal data linked across programs could best address these questions.

Feedback for the FSS Data Center

Better Use of Data

Interviewees were asked general questions about how data could be better used in FSS policy, research, and administration; their answers reinforce the value of the Data Center's mission. On the whole, respondents stressed the importance of increasing the accessibility and usability of data and of improved communication between administrators, researchers, and policy makers. These are important key goals for the Data Center.

The Role of the Data Center

Nearly all interviewees agreed the FSS Data Center could play a crucial role in helping them achieve their research and data needs, particularly by providing access to data from both other agencies and other states they cannot easily access.

Interviewees also see a role for the Data Center in helping to use, analyze, and translate their data more efficiently. Interviewees feel the Data Center can help provide the tools needed for more sophisticated data trend analysis, analytics, and decision making support on the ground, which would also help translate data to inform practice.

Technical Assistance, Training, and Peer Engagement

Most interviewees noted that a critical function of the Data Center will be to provide training and technical assistance (TA) to states and data users. Suggested training topics include statistical methodology, measurement issues, linking data, program evaluation, research methods, and how to manage and interpret data. Additional training and TA topics suggested included how to compare data across states, addressing data quality concerns, the practical use of the Data Center, and more personalized TA to states to answer specific questions or topics of interest. Many interviewees also suggested the Data Center can encourage peer engagement among data users.

Potential Implementation Challenges

Although the Data Center was almost universally seen as a potentially helpful resource for accessing, using, translating, and analyzing data, many interviewees cautioned us about challenges they anticipated for the implementation of the center. These largely revolved around getting states to participate, mainly due to legal, security, and political reasons, but interviewees also expressed concerns about data quality and comparability.

Translating Findings into Actionable Steps

The needs assessment findings suggest a number of ways in which the Data Center could support and facilitate FSS research. The many tasks and roles of the Data Center fall into the following general categories:

- **Data access and permissions.** The Data Center will assist with negotiating and establishing data sharing agreements.
- **Scope and types of FSS data.** The Data Center will help agencies access datasets from other program and topic areas to contextualize their work. Among other tasks, the Data Center can provide a menu of the scope and types of FSS data that are available in FSS research.
- **Data security, storage, and documentation.**
- **Data management, linking, and analysis.**
- **Training, technical assistance, and peer engagement.**

Next Steps

With the needs assessment complete, the project team, in consultation with OPRE, will continue to develop the framework for the Data Center while broadening discussions with states around their involvement in the Data Center. The Data Center will undergo a pilot year, which will be crucial to testing the center's capabilities and troubleshooting glitches. Throughout and following the pilot year, the Data Center will acquire datasets (including state and local administrative data, national and original survey data, and evaluation archives) and then become fully operational, all the while providing technical assistance to data users and providers.

Introduction

Family Self-Sufficiency Data Center

The Family Self-Sufficiency (FSS) Data Center is funded by the U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. The Data Center's primary goal is to enhance the use of administrative and survey data to answer policy-relevant questions surrounding family self-sufficiency and well-being. The Data Center pursues this goal by working to increase the amount and quality of data available for FSS research. Along with the primary focus of state and local administrative records, the Data Center will also seek out, bring together, and regularly update relevant surveys and program evaluation efforts, allowing data users to access the data they need and to compare findings across difference sources and types of data. In addition to making existing and new data sources easier to access, the Data Center will work with policymakers and researchers to ensure data can be stored and analyzed in a secure environment. The Data Center will provide an empirical, data-driven research platform for researchers, policymakers, and administrators to use in answering fundamental policy and program questions. It will build knowledge that will ultimately be translated into better policy and practice.

The FSS Data Center was launched in 2013 and has focused on several initial activities necessary to build and test a secure data environment. These have ranged from beginning outreach and collaboration efforts with potential Data Center providers and users to pilot testing an initial version of the platform. Of particular importance in the Data Center's early activities has been a needs assessment conducted with key stakeholders and the development of the technical framework for the Data Center. This report provides a summary of the needs assessment process and findings from that process, which will inform the other activities of the Data Center.

Needs Assessment Overview

The objective of the needs assessment is to obtain input from key stakeholders—primarily from potential Data Center users—regarding their current capacity, challenges, and needs with FSS data in order to learn how the Data Center can help meet those needs. Specifically, the needs assessment solicited information in three areas: (1) the topics included in FSS, (2) the current use of data and research about FSS, and (3) the data and assistance needed among the FSS data user community.

Topics Included in Family Self-Sufficiency

Given the range of topics that can be considered part of the conceptualization of FSS, the needs assessment asked questions of informants about their perspective of FSS, including what programs and populations should be included when considering the scope of FSS for the Data Center. FSS has many components. For example, the Self-Sufficiency Research Clearinghouse (SSRC) articulated twelve self-sufficiency topic areas: asset building, tax policies, and subsidies; child care; child support; community development and housing; education and training; employment; family formation and family structure; food assistance; general research on income and poverty; health; TANF policy, services, and benefits; and transportation. FSS may be focused on specific policies or programs or more general goals for society. Understanding how potential users perceive FSS is important in determining what areas of FSS the Data Center should prioritize and who should be engaged in the development of the Center.

Current Use of Data and Research

Knowing the Data Center must address data needs and analysis goals across the entire spectrum of potential data users, the needs assessment solicited input from interviewees about their experience using FSS data and their research efforts. The Data Center is intended to augment the variety of ways data is currently used by policymakers, researchers, and analysts to analyze and understand self-sufficiency. The needs assessment discussed this variety of uses and needs of data and research around FSS with the Center's potential clients in order to learn how the Data Center can enhance clients' abilities to conduct research. These discussions also touched upon factors that might impact public agencies' ability and willingness to participate in the Data Center, including their capacity to analyze data and their challenges with accessing and using data.

Data and Assistance Needs

Given the broad scope of the concept of FSS and the range of experiences of current data users, the needs assessment inquired about specific data needed for FSS research as well as what technical assistance could be useful to support greater use of FSS data. Similarly, because many users will require training

and support to use these data effectively, interviewee feedback was solicited to help shape the plan for technical assistance with future Data Center users.

Methods

To achieve the objective of the needs assessment, the team conducted interviews and focus groups with key FSS stakeholders, including program administrators, researchers, and other interested groups, such as professional associations and advocacy groups. Most interviews were conducted via telephone, while the focus groups were conducted in person. The interview and focus group questions were tailored to the specific type of respondent to ensure the most appropriate questions were asked of each individual based on their role and use of data. The needs assessment planning began in January 2014; focus groups and interviews were conducted between February and May 2014.

Sample

Using convenience and snowball sampling, a purposeful sample was obtained to ensure a range of stakeholders were represented. The sample included public agency administrators at the federal, state, and local levels. State and local public agency administrators were selected to reflect a range of urban and rural areas. While program directors and managers of the primary self-sufficiency programs (TANF and SNAP) were targeted more heavily in the sample, other programs related to self-sufficiency were also included (for example, vocational rehabilitation, workforce services, and others.). The sample also included researchers and other data users, such as foundations and advocacy groups. Individuals were identified for the sample based on recommendations from the project team and other interviewees.

A total of 96 people participated in either a focus group or interview. The majority of sample participants were administrators from public agencies (61 individuals, comprising 63.5% of the sample). Researchers represented about a quarter of the sample (23 individuals, comprising 24%), while other stakeholders represented the smallest group (12 individuals, comprising 12.5% of the sample).

Focus Groups

Two focus groups were held in February 2014. One focus group was composed of members of the University of Chicago Advisory Council for the FSS Data Center project. This group included professors with expertise in FSS research from numerous disciplines, such as social work, public policy, and economics. Another focus group was held with the FSS Scholars, a select group of FSS researchers who are also supported by the U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. In total, 15 researchers participated in these focus groups— 8 from the Advisory Council and 7 FSS Scholars.

Interviews

The majority of the needs assessment was composed of interviews with federal, state, and local agency administrators, along with researchers and other stakeholders. A total of 81 people participated in 45 interviews. The largest group of participants was from state public agencies; 53 people from 23 states were interviewed. The states represented were regionally and economically diverse, incorporated both urban and rural populations, and included both state- and county-administered systems.¹ Seven local public agency administrators from four local areas were also interviewed.² In addition to state- and local-level public agency representatives, one federal-level administrator was interviewed. Eight researchers from eight academic or research institutions were also interviewed because of their expertise with FSS research and data.³ Finally, there were 12 interviewees from 6 professional or advocacy organizations with a focus on FSS.⁴

Analysis

Once the needs assessment focus groups and interviews were completed, notes were qualitatively analyzed using ATLAS.ti. Given the subjective natures of qualitative analysis, each set of notes were coded and analyzed by two researchers, using preliminary interrater reliability checks to promote consistency and accuracy with the analysis. The results of the analysis are described in the next sections of this report.

¹ Public agency representatives from the following states were included in the sample: AL, AZ, CA, CO, FL, HI, IA, MD, MI, MN, NC, NH, NV, OH, OK, OR, PA, SC, TN, UT, VA, WA, and WI.

² Local agency representatives from the following local areas were included in the sample: Washington, DC; New York, NY; Los Angeles, CA; and Allegheny County, PA.

³ Researchers from the following academic or research institutions were included in the sample: City University of New York, University of Pennsylvania, University of Chicago, University of North Carolina at Chapel Hill, University of Wisconsin at Madison, Mathematica, Upjohn Institute, and Urban Institute.

⁴ Representatives from the following professional or advocacy organizations were included in the sample: American Public Human Services Association, American Enterprise Institute, Center for Budget and Policy Priorities, Children's Data Network, National Governor's Association, and the Self Sufficiency Research Clearinghouse.

Concept of Family Self-Sufficiency

Different interview respondents defined family self-sufficiency in different ways. The needs assessment attempted to capture a complete picture of the diversity around the concept of FSS that is in use among administrators, researchers, and other data users, as well as explore the interviewees' notions of how FSS can be shaped by local and familial context.

Defining Family Self-Sufficiency

In its narrowest definition, FSS involves families who cannot support themselves because adult members are unemployed, underemployed, or have a disability, or the family includes a child with a disability that prevents one or more adults from being fully employed. The programs that were most frequently referenced as promoting FSS were programs that either enabled family members to work or supplemented family income in the absence of living wage employment. A majority of interviewees cited TANF, SNAP and other food assistance programs, unemployment insurance, and child care subsidies as key FSS programs. Other commonly mentioned services and programs included housing (public housing, Section 8 vouchers and other housing support, and homelessness services) and job training. Tax credits can provide another support for struggling families; one interviewee noted that the Earned Income Tax Credit (EITC) has the largest impact on FSS of any public program. A few interviewees cited other, smaller public assistance programs, such as energy assistance and refugee or immigrant services.

Many interviewees broadened their definition of FSS to embrace dimensions beyond employment and income that impact the stability, health, and well-being of the family. Interviewees frequently mentioned the importance of Medicaid and health services, including free or reduced price medical and mental health care. (As several respondents noted, medical problems can significantly constraint a person's ability to function independently.) Similarly, education and child welfare were cited by several interviewees as elements of FSS. Although they do not relate directly to the family's financial health, they are important components of general well-being. Other topics mentioned related to contextual challenges affecting the

family's circumstances, including criminal justice, domestic violence and substance abuse, economic and community development, transportation, and assets. For example, one interviewee explained that assets are frequently overlooked but can determine whether a family can remain self-sufficient through an unexpected cost, like car repairs.

In general, interviewees believed that understanding a household or individual's context allows analysts, case managers, or researchers to better understand what may be shaping observed behaviors. It also allows for more efficiently targeting programs to maximize their impact. As one researcher explained, "If we know more about their particular circumstance, we can help them. . . sooner." Location-related contextual factors, like cost of living or community strength and resources, can significantly impact FSS. Families may also grapple with issues like substance abuse or domestic violence that make self-sufficiency harder to achieve or sustain.

Respondents emphasized that FSS is best viewed as progress along a continuum, rather than something to be attained. "People say. . . 'reaching self-sufficiency'. . . like you cross some line and then you're self-sufficient," observed one researcher. Interviewees who discussed this idea often pointed out that just because families do not rely on public assistance does not mean they are self-sufficient. It is important to remember eligible nonrecipients as well as individuals or families who are not eligible for programs but still have need, especially those earning just slightly more than would make them eligible for assistance. Because research on FSS often utilizes administrative data from public programs, families that do not receive these benefits are difficult to study and often overlooked.

Similarly, families may increase their self-sufficiency through the use of public benefits. For example, for most families moving out of public housing would represent a step toward FSS. However, for a homeless family, moving *into* public housing is a sign of increased self-sufficiency. Other individuals and families face physical or mental health challenges that may make complete independence infeasible; in these cases, interviewees spoke about the need to achieve the "highest [possible] level of self-sufficiency." Respondents emphasized the particular challenges of high needs populations, including immigrants and refugees, the homeless, the extremely poor, the formerly incarcerated, teens coming out of foster care, veterans, and families utilizing multiple programs. In these cases, it is difficult to decide what FSS is and what constitutes successful results for an FSS program.

One interviewee described self-sufficiency not as a lack of benefits but as a sense of empowerment; the self-sufficient family has "the capacity to case manage themselves" and to "enjoy a decent standard of living." By this measure, FSS is a matter of well-being rather than autonomy from public benefits. This approach also makes it possible to view families as self-sufficient in some ways and at some times but not others.

Several interviewees questioned whether the phrase “family self-sufficiency” adequately captured these complexities and whether the phrase is potentially stigmatizing to poor families. A few administrators mentioned that their programs began using the term during the welfare reform era; one researcher also described it as “left over from welfare reform jargon.” These respondents were concerned that “self-sufficiency” has been conflated with receipt of a particular set of means-tested public benefits. However, other public benefits—such as the K–12 education system, Medicare, and tax credits—were mentioned as examples of benefits received by middle and upper income families that are often excluded from the FSS umbrella. It is socially acceptable for “self-sufficient” families to receive those benefits. Several of these interviewees said that they preferred to talk about “family well-being”; they felt that phrase was less political and more naturally implied the full slate of services and circumstances that create stable and healthy families.

The Data Center will include a broad array of data—both in scope and type—for its users to access. This is in response both to the broad catalog of topics interviewees included in their definitions and the concerns that research should not focus exclusively on economic stability but should think about family well-being as a whole.

Overview of the Use of Family Self-Sufficiency Data

This section describes how users, especially state and local administrators, are currently using data for FSS work, and how capacity, data sharing, and other challenges impact this use. This context for FSS data use informs how the Data Center can best complement administrators' and researchers' current work.

FSS Data Used for Reporting and Performance Management

Agency administrators frequently discussed examples of using data for reporting, performance management, and quality control purposes.

A number of administrators mentioned reporting as a primary use of data. In addition to federally mandated reporting, many agencies use data to comply with requirements at the state level and to conduct ad hoc reporting to address particular questions of interest. Several administrators noted that the data needed for mandated reports was generally some of the cleanest data available and that the metrics tracked in these reports—the TANF work participation rate, adoption numbers, the food stamp error rate, and other metrics—were a key focus of performance management efforts. One interviewee asserted that his department “live[s] and breathe[s] those data elements.” In some cases, federal reporting also made metrics from other states available for comparison purposes, but several states noted that both the availability and value of this data for benchmarking were extremely limited. One respondent explained that while it is possible to see which states are performing better or worse, it isn't possible to know why at a glance, or to compare the same programs or populations across states.

In addition to external reporting, almost half of administrators interviewed discussed regularly using some kind of metrics to monitor performance. These metrics are generally aggregated into scorecards, dashboards, or other formats and used to compare results across departments, geographies, or service

providers. Some of the data points cited include take-up rates (what percent of eligible families use a program), recidivism rates (the rate at which clients who leave return to the program), the volume of calls and intakes to the department, and the timeliness of intake processing. State and local departments aim to use these data points to inform and adjust practice. For example, an office may adjust the number of staff working at particular times to reflect the volume of visits, or it may update information online to reduce the need for those visits in the first place. One department started tracking engagement and noticed that the rates at which clients showed up for an orientation meeting after their initial eligibility screening was extremely low. The department subsequently made orientation a requirement for eligibility and this rate improved.

Departments also compare their client data with data sources describing the local population to get a sense for who is and is not being reached by programs. For example, a comparison of service utilization to poverty rate across locations helps administrators understand which communities may be over- or underserved. The same kind of analysis can demonstrate whether certain demographic groups are more or less likely to participate in programs or highlight populations that utilize resources at a particularly high rate. In turn, these analyses can inform outreach efforts.

A few administrators mentioned that their departments are trying to move to metrics that reflect family well-being more broadly in order to encourage staff members to think holistically about the good of the family, rather than focusing only on particular aspects or individuals. For example, some states are introducing two-generation initiatives where they consider not only a mother's work outcomes but also the outcomes of her children. This kind of approach necessitates using more extensive data linked across multiple data systems to develop meaningful benchmarks or assessments.

It was relatively common for administrators to also discuss using data for fraud prevention and program integrity. Some interviewees expressed interest in getting access to death and incarceration records specifically to bolster these quality control efforts. Others applied predictive analytics techniques to model instances of fraud.

Although agency administrators were much more likely to use FSS data for performance management and quality control purposes, the questions they asked require access to the same kind of data researchers seek. This is particularly important for administrators who want data to understand program impact. Administrators mentioned several outstanding performance management questions, including:

- What is the ideal caseload size?
- How can we improve program take-up rates among eligible households or individuals?
- Has participation in the program improved participants' outcomes?

- How do performance management policies impact street-level caseworker practice and interactions with clients?
- When benchmarks are attained or program trends move in the intended direction, can analysts identify the factors driving the changes?

Being able to answer these questions requires access to data connected across multiple sources and over a series of time points. Researchers and administrators shared the same processes and challenges in accessing and using this data.

FSS Data Used for Research

In addition to traditional policy research, FSS data was used by a small group of administrators to answer research questions that would inform policy planning. For example, one administrator talked about a project where staff drew on Census data and data from other departments to understand what kind of resources were available in each community. This allowed administrators to contextualize performance across locations. In another case, the agency sought to model the impact of changes on the population, including the impact of the Affordable Care Act or the impact of a change in the minimum wage. In general, where agencies talked about using data for research, it was usually driven by a desire to understand the impact of a policy that had been enacted or was under consideration.

Almost one-third of state and local administrators interviewed mentioned undertaking partnerships with external researchers or organizations to address FSS questions. These partnerships focused on evaluating programs, understanding populations, and estimating policy impacts—similar to the FSS-oriented work of the researchers interviewed. The Data Center has the opportunity to serve as a relationship broker to foster such partnerships between administrators, researchers, and other data users, allowing them to fulfill research needs.

Capacity and Challenges Using FSS Data

How agencies and researchers use data is determined not only by what they seek to know but, perhaps more importantly, by their ability to access and analyze the necessary data. The experiences of our interviewees highlight gaps in data usability that provide potential opportunities for the Data Center. This section summarizes interviewees' current capacity to use data, their ability to share their organization's data, and other challenges and limitations they have faced while using data for FSS work.

Agency Staff Capacity to Use Data

State and local agencies' capacity to use data varied dramatically, as described by interviewees. About one-third of administrators interviewed described having dedicated research or evaluation staff members within their agency—often small teams focused on addressing research questions, developing databases,

and developing performance measures related to data. These teams are armed with the tools to use advanced statistical techniques to evaluate outcomes to determine program impact.

Another one-third of administrators came from agencies with no formal research staff, but where one or more staff members are dedicated to data management and reporting. Although these agencies may not be able to glean the same kind of insights from their data or do not have the support of a Research and Evaluation team, they do have the capacity to get at basic program metrics and manage data extracts. One state recently hired a full-time data manager to create the database and help link data; another mentioned hiring someone to help with performance management-type projects in house so they don't need to be farmed out to another department or team.

A handful of interviewees described working in agencies that lack even that capacity. In these agencies, program staff and administrators do their own reporting or data analysis, depending on their time, skill, and inclination. Data management and reporting in these agencies often falls on the IT staff. These IT teams are often supporting multiple departments and may have little understanding of agency programs and operations.

As previously mentioned, a number of states enter into partnerships with external researchers in order to address some of their research needs. These partnerships also increase their capacity to do research and evaluation work. They require data sharing agreements; the process of arranging these agreements varies in complexity on a state-by-state and, sometimes, an agency-by-agency basis. The process also requires staff expertise and resources, which will be discussed in the next section. One state administrator acknowledged that their state has contracted with a local university to do their data and analysis work. Even when responding to simple requests from the General Assembly, this person noted, state employees cannot produce an answer internally and send the request to the university researchers for a quick response.

However, even if a state or locality has favorable data sharing policies and wishes to participate in a research project, if the agency lacks the basic capacity to pull data extracts and help researchers understand the context and quality of data, it can be difficult for the state to take advantage of these kinds of partnerships. One researcher described this as a “Catch-22”—most policy research is done in the states with the most generous benefits, as they typically have the largest human services budgets and have the resources to extract the richest datasets for research. As a result, the smaller, leaner human services programs don't have the resources to extract data for research and are not often the subject of research projects. Good quality data costs money and can take a significant portion of the budget—money that is pulled away from direct service. The cost of working with external researchers such as those at a local university can be “prohibitively expensive” for some smaller states.

Several interviewees discussed having tighter budgets or reduced staff resources for research in recent years, especially in the wake of the 2008 recession. These agencies reduced their research or data staff or eliminated units dedicated to working with data. One interviewee whose team usually had a \$1 million budget stated simply, “In 2008, my budget went to \$0.” His boss told him, “It’s between research or helping homeless people.” However, in that particular case, funding for fraud prevention work was used to keep staff and the budget was fully restored by 2010. Several other interviewees also mentioned that funding had been restored or, at a minimum, that reductions were shared across teams; research and data staff were rarely the targets of budget cuts.

Almost as many interviewees, however, described staff capacity as having *increased* in recent years. “We were probably the only state hiring during the recession,” one administrator commented. Others mentioned that the emphasis on maximizing the agency’s impact despite having a reduced budget led to a renewed focus on data and research. “The harder times are, the more it calls out for making that investment,” and maximizing the use of diminishing resources. Other interviewees noted that the recent recession did not impact their capacity directly, since despite layoffs or hiring freezes within the agency, their data and analysis capacity was not affected.

Outside of budget changes, interviewees also cited skills gaps and the demand for analysis as impacting staff capacity. A couple interviewees described challenges recruiting and retaining qualified staff. This was especially a challenge for states with legacy data systems; as staff members trained on these systems retire, it can be difficult to find young staff with the necessary experience in dated systems. At the same time, as agency and state leadership increasingly emphasize data and metrics, departments may find themselves doing more with the same number of staff. One administrator asserted that the motto of her state’s government is “increased demand, decreased capacity.”

When asked about capacity, interviewees also often talked about the ways in which available technology increased or decreased capacity. New, updated data systems can expand capacity and make research or analysis easier. Several interviewees mentioned use of technology—such as iPads for real-time data entry for case managers—or system improvements that made their data more available and easier to use. A few others described their old systems as being a particular limitation on capacity. One administrator, from an agency where it was very difficult to extract data from a legacy system, complained, “It takes twenty people to do our data analysis.” Projects to replace old systems also consumed staff time and resources, however, further decreasing capacity in the short term. The Data Center hopes to make data more accessible and easier to analyze with user friendly analytic and reporting tools. A challenge for the Data Center will be to obtain the data in the first place from the states with limited capacity and resources to extract and work with their data.

Agency Data Use Capacity

In addition to staff and resource capacity, interviewees discussed other impacts on their capacity to share, receive, and use data. Administrators talked mostly about data sharing, especially sharing of individual-level data, in two contexts: sharing with external researchers, and sharing with other agencies within the same state. Interviewees occasionally mentioned sharing data with journalists—some administrators asserted that journalists represented the majority of their data requests—but these requests were generally for published or otherwise aggregated data. No interviewees discussed sharing data across states, although a few mentioned state and local partnerships in their area. Most administrators exclusively talked about sharing their own agency’s administrative data, but in a few cases where interviewees worked with state data hubs or other integrated data systems, they discussed processes for sharing data that they had received from another department. Additionally, the capacity to access data from other departments and agencies varied widely across interviewees. There are several factors identified that can promote data sharing and access, as well as a few that act as an impediment to an administrator’s ability to share and access data.

Promote Data Sharing and Access: Clear data sharing policies

Where states and municipalities had clear data sharing policies, they were generally similar, requiring data sharing agreements, evidence of the recipient’s ability to securely store the data, and IRB approval for any research. Most agencies made it a policy to only share de-identified data, although they were often willing to match data from multiple department sources and provide the data linked and stripped of identifiers. One researcher noted that public offices generally became much more willing to consider sharing data if the researcher can agree from the start that social security numbers will not be part of the request. Where individually identified data is required—as for evaluating specific programs or identifying individuals to target for other interventions—many departments have policies that require collecting informed consent from participants.

Promote Data Sharing and Access: Collaborations and relationship building

Where administrators and researchers described successful collaborations and data sharing arrangements, they repeatedly emphasized the importance of relationships in making these arrangements work. Organizations which work together frequently and perform mutually beneficial work tend to have less trouble negotiating data sharing agreements. It is crucial that external researchers earn and maintain the trust of agencies with which they collaborate. As researchers learn the state’s data, the capacity burden of data sharing with that individual is lessened. Several researchers noted that they had worked hard to build relationships and found it much more difficult to request data outside of the agencies where they had developed those relationships.

Impede Data Sharing and Access: Legal restrictions

However, many administrators found data sharing extremely difficult or uncommon in practice; oftentimes, data sharing policies in these states are unclear or, if clear, they are extremely stringent. Tight restrictions on data sharing stem from the possibility that data could be linked to other public data, thus identifying individuals. This creates questions about when data is actually confidential. One interviewee explained, “When you have to open the door about an agreement [i.e., when the data includes more information than would be permissible for public use], the threshold leaps 10,000 feet.” In one state, even data sharing agreements between departments require approval at the level of the Secretary’s office on both sides, a process that is only practical for very large grants. Another administrator said that agency heads have legal permission to share data if doing so is “in the furtherance of [the commissioner’s] responsibilities,” a requirement department leadership interpreted narrowly. On the receiving side of the process, researchers noted that state interpretations of what could be shared are idiosyncratic and political.

Time limits on agreements can pose additional barriers. Reaching agreements can be a lengthy process, and sometimes the time period for which an agreement is valid can be brief. This can lead to issues of the data quality (when one finally achieves permission to use the data, the data are outdated) or to issues for the data user (once the user has cleaned up the data, the time limit expires and the data must be returned).

Both researchers and administrators mentioned a desire for help finding clarity about data sharing practices at a national level. One researcher suggested that it would be helpful for the Data Center to provide sample data sharing agreements or protocols that researchers could reference. Another interviewee had heard that U.S. Department of Health and Human Services is creating a “confidentiality toolkit” and is looking forward to having this resource.

Accessing data sets from other departments was a significant challenge for all interviewees. Some data systems appear more challenging to access than others, often due to privacy or legal concerns. Education, wage, mental health, and child welfare data often require additional confidentiality provisions that other public benefit data do not. These access restrictions can, at times, keep administrators from accessing the data altogether. Several states mentioned, in particular, the restrictive nature of federal data. Some agencies (the IRS, for example) require specific confidentiality training or special access arrangements, such as users remaining onsite while using the data. Furthermore, once data is obtained, linking that data can be challenging due to restrictions on accessing identifying information. For example, interviewees noted particular challenges linking survey or Census files to administrative data or linking education data with other data.

Interviewees faced many of the same legal challenges and confusions in accessing data that they wrestled with when trying to share data. Several interviewees echoed the difficulty in obtaining permissions,

working through data sharing agreements, and obtaining consent to access data from other agencies. One interviewee referred to the access challenges merely as “bureaucratic foolishness” with the governing agency that owned the data making data access an unnecessarily lengthy and tedious process. Another interviewee noted that departments want a legislative mandate that formally gives them permission to hand over their data. One interviewee remarked that, recently, data security and confidentiality became more time-consuming and complicated. This was largely due to agency leaders witnessing data breaches.

Impede Data Sharing and Access: Resource limitations

Resource limitations, such as limited staff or limited system capacity, also impacted data sharing. As previously mentioned, limited staff capacity to work with data could make it hard for states to prepare requested extracts. Even when agencies had the staff capacity to manage data, several administrators noted that staff time is needed to answer questions from researchers who lack knowledge of the agency or its data. Time dedicated to answering research questions and performing analysis and evaluation tasks is often time taken away from direct service, especially in agencies where there is no dedicated research and evaluation team. Additionally, many administrators are working with outdated data systems that make data extraction and analysis a tedious job. For these reasons, it is especially important that researchers perform work that benefits the agency.

Impede Data Sharing and Access: Data not collected

Another challenge interviewees mentioned was that some data were not as helpful as anticipated. For example, given that so much data is now being collected for programs, some of the data may not be useful—perhaps the data is too general and not specific enough for the purpose intended. Given that data received from some data systems are meant to manage programs and not intended for research purposes, there is a need to translate program data into something that can be used for research before it can be beneficial. Lastly, three interviewees discussed elements they would like to have access to but are not collected and stored anywhere. These elements included untaxed wages and the history of clients’ domestic abuse incidents.

Data by Type

During the needs assessment, interviewers engaged administrators and researchers in detailed discussions about the data they use and have access to in their work, as well as the data they want but do not have available or use to the extent they would like. Many of the general challenges, quality concerns, and issues around using the data will be discussed in other sections of the paper. This section details the specifics of the types of data interviewees most commonly use and would like to use in their line of work. It begins with a discussion of longitudinal and linked data generally and then focuses on particular data types, organized by administrative program or data collection method. The following types of data will be discussed: TANF; SNAP and other food assistance; unemployment insurance (UI) earnings and other employment data; Medicaid and other health services; education, child care subsidies, and child support; child welfare; criminal justice; tax and EITC data; other supports including housing, transportation, and SSI/SSDI; Census Bureau data; and survey and evaluation data.

Longitudinal and Linked Data

Many interviewees emphasized the importance of both longitudinal and linked data in analyzing and researching FSS topics. Linked data involved connecting data sets from different agencies, programs, or sources other than agencies. These data may be longitudinal in nature, but often may be cross-sectional. Access to more and different linked data was a top priority for almost everyone interviewed. Additionally, many of the questions that both agency administrators and researchers hoped to answer based on FSS data require access to longitudinal data and the capacity to work with and analyze such data.

Longitudinal Data

About half of interviewees had access to a data warehouse or another source of longitudinal data to draw comparisons and track participants over time. A couple of interviewees wished they could also access longitudinal data for other states to follow participants and outcomes across state boundaries. Several others wanted to be able to follow participants chronologically after they leave the system altogether.

Included in this group are several states who expressed a desire to follow TANF leavers and understand their employment retention and wages as well as uptake of other public benefits down the road.

Linked Data

About half of interviewees worked with some kind of integrated database—a database, data hub, or data warehouse—in their agency that integrates multiple data sources. All included at least SNAP and TANF data in one database; most integrated other human service data as well, including Medicaid eligibility, child welfare, child care subsidies, child support, and sometimes housing data. In a few cases earnings and employment, criminal justice, or education data were either included or could be linked, in an ad hoc database, with the existing database. For example, one state administrator discussed the Integrated Client Database (ICD) which was specific to his agency’s programs and contained core linked data. However, the agency had the ability to run dozens of additional linkage processes to bring in ad hoc populations, such as those in Department of Labor programs, and link them with the ICD. A handful of other interviewees mentioned that these kinds of integrated systems were being developed, though that process was not without challenges. One administrator described the experience of “private sector consulting companies circling around you.” According to this interviewee, the consulting companies would say, “If you just pay us \$10 million, we’ll put that all together.” Another administrator discussed the process currently underway in the state to broker a relationship between the Census Bureau and two human service departments in the state in order to link survey and administrative data. The administrator acknowledged there are lengthy delays because of issues with the data sharing agreements.

The presence of linked data is often tied to demands of reporting requirements, regulations, or program benchmarks. For example, the data most frequently linked on an ad hoc basis were TANF data with employment or earnings data. These data were usually linked with an eye toward verifying eligibility. Other human service data, as well as criminal justice and education data, were occasionally linked. However, outside of integrated data systems it was common for only two files to be linked at a time.

There are many ways to link records over time or to link individuals across programs. In most cases social security numbers or another internal identifier were used as common identifiers, but addresses could be used in certain circumstances as well. Several interviewees mentioned that data quality concerns with identifiers adversely impacted the quality of the linkage and resulting data. Some interviewees used probabilistic record linkage⁵ instead of relying solely on common identifiers.

⁵ Probabilistic record linkage (PRL) relies on defined probabilities to determine whether two records describe the same individual. PRL uses the probability that two records will agree or disagree on a given field (for example, first name or social security number) if the records represent the same entity, and the probability that the records will agree or disagree on that field if

Many respondents expressed the need for greater access to linked or longitudinal data. In some instances, the focus was on pulling in data from other programs that may be difficult to obtain, for reasons discussed previously, such as criminal justice or EITC data. Additionally, many respondents discussed the difficulty they had linking with other data sets that use identifiers different from their own. Specifically, education and child welfare data were often cited as datasets that would be valuable to link data with, but many challenges arise because standard social security number (SSN) identifiers are not typically used in these two datasets. Respondents also sought to link federal data, like Supplemental Security Income (SSI) and Social Security Disability Insurance (SSDI), to the records of state benefit recipients. However, due to the restrictive nature of various federal data systems, this is often not possible.

Ultimately, many respondents wanted to better understand a family's full circumstances by integrating as much data as possible about the family's needs and resources. This "holistic" view of the family captures the context respondents sought in understanding self-sufficiency and well-being at the family level. Furthermore, better integration of human service data facilitates the "no wrong door" approach to case management, where a single caseworker can assist clients with an array of services. This streamlines caseworker efforts and enhances the efficiencies of the network of public benefit programs serving an individual.

Data by Source

TANF Data

As a result of our sampling approach, which focused on TANF administrators and researchers with experience studying welfare programs, TANF data was the most commonly cited data source used by interviewees. Almost all interviewees from state or local agencies used TANF data or worked in a department with individuals who used TANF data. About half of the researchers and others interviewed also mentioned using TANF data.

Administrators used TANF data primarily for mandated reporting and to track performance management metrics. Data about job placement and work activity participation were mentioned frequently, since many departments use these numbers to assess the effectiveness of their programs as work supports. One

they represent different entities. These probabilities are aggregated across fields, allowing the analyst to make an informed decision about the overall likelihood that the records match.

administrator gave the example of a typical month's report to the state governor, which included work participation rates, total work hours, job placement information, individuals placed in jobs paying more than \$10 per hour, and detail on individuals' time spent not doing work activities. Interviewees were also interested in TANF leavers, and, in addition to the administrative data, several surveyed departing clients about their circumstances, including wages, hours, and industries in which they planned to work. This information allows administrators to see the kinds of jobs their clients are taking, and adds to data that enhances understanding about clients who return to TANF.

Since most public agency interviewees oversaw TANF programs, many of the data sharing experiences discussed previously were relevant with TANF data. States' willingness to share this data varies widely, as do their legal requirements and technical capabilities. The quality and extent of data on TANF clients is variable, and most interviewees mentioned their wish for more or better data. The data that is frequently tracked and used for key metrics, like the work participation rate, is generally of good quality. However, other data often suffers from hurried data entry or missing data fields. Most demographic data and data about individuals' circumstances before and after TANF receipt are available only as self-reports, which also comes with quality concerns.

Administrators often viewed linking TANF data to other sources as a way to address these limitations. They hoped to access data about clients' education, incarceration history, receipt of other public benefits, and struggles with issues such as domestic violence or substance abuse. This would allow them to provide a more comprehensive picture of the resources available to clients and the barriers they face. In a few states, comprehensive data hubs provided these contextual data for analysis. In many states, however, the extent of available integrated data focused primarily on participation and eligibility with TANF, SNAP, and sometimes Medicaid or other state benefit programs. Integrating these data sources potentially increases the overall quality and extent of the data. It also allows program administrators to look at instances where eligible individuals do not claim all the benefits they could. However, these data systems do not include all of the data sources that interviewees wish to integrate.

In addition to linked data, respondents wanted access to more longitudinal TANF data to examine postprogram participant outcomes. They also wanted data from other jurisdictions for benchmarking purposes. As most interviewees have access to TANF data, they are especially aware of the limitations of point-in-time administrative data in answering longitudinal questions about client outcomes. Interviewees also mentioned wanting to compare their TANF data against other states to benchmark outcomes.

Food Assistance Data

SNAP data was used nearly as frequently as TANF data, in part because administrators and researchers we spoke with noted they generally do work on both programs. SNAP data was used by over 80% of interviewed administrators and about one-third of researchers and others interviewed.

However, interviewees gave relatively few examples of using SNAP data specifically. They often spoke of SNAP and TANF data as a unit, since many states appear to track eligibility data for these two programs in a common database. One administrator mentioned that the SNAP error rate was of particular interest for his department. Another mentioned using data on SNAP work programs and work activities. Although no administrators spoke about experiences sharing SNAP data specifically, two researchers noted that SNAP data can be difficult for outside parties to access. They cited legal restrictions on this data and a lack of understanding at the state level of how the data may be used.

Interviewees infrequently spoke about specific challenges and desires with regard to SNAP data. However, in general, as with TANF data, respondents were interested in seeing participation data combined with other data sets that can provide information about participants. Interviewees were also interested in what percent of the eligible population actually enrolled in SNAP benefits, a question that can be addressed by comparing SNAP enrollment against enrollment in other public programs or by comparing SNAP enrollment numbers against general poverty statistics from public data sources. Interviewees also want to see SNAP data available for other geographies, including data at the county or national level.

A few interviewees mentioned wanting to have access to data for food assistance programs other than SNAP, including state-level programs and the WIC program data. A few respondents specifically noted they would like to use the WIC data from their state but don't currently have access to the data.

Unemployment Insurance, Earnings, and Employment Data

The majority of interviewees—both state and local administrators, as well as researchers and other data users—acknowledged using some form of data related to employment and wages. The interviewees had a wide range of types of access to both UI earnings and other wage and employment data. Some had full access to UI earnings data, while others only had partial access and some had no access at all. Some use Social Security data as their source of wage data, which is quarterly and contains earning histories that are part of calculating SSI benefits. Some use data from various departments within their state (such as the Departments of Labor, Economic Development and Tourism, Employment and Training, and other departments) to gather data related to employment or aggregate-level data. Another mentioned using the work number database for income eligibility verification for benefits, but that is point-in-time data and its

usefulness for analysis is limited.⁶ The same interviewee admitted most other employment information comes from the centralized assessment tool that asks for past employment information. Finally, a few state agency interviewees mentioned using the National Database of New Hires (NDNH)⁷ as an ad hoc way to obtain employment data. Researchers also discussed their access to employment data. One researcher described having access to large sets of state data, with the UI and employment security agencies linked, thus providing a massive data system with a wealth of data available for research.

Agencies have many different kinds of access to data, and interviewees displayed a range of sophistication in how they use employment data. Some used earnings and employment data to track individuals across time and to track their participation in work activities. One interviewee noted being able to track people who are leaving programs with wages, including the amount of those wages, as well as employment information, such as number of hours worked and industry of employment. They wanted this data to know which activities are most effective in achieving increased work participation rates. Another interviewee explained that they can track whether an individual attends a job fair and becomes employed as a result.

As expected, given the range of use and access among the state administrators interviewed, there is also wide variation in the level of data sharing and the ability to link employment data with other systems within the state. However, there are several examples of states successfully sharing and linking their UI earnings data with other programs, most notably with TANF and SNAP. One state interviewee acknowledged that due to the focus on getting people to record the work status of TANF recipients, state UI earnings data from the Department of Labor are frequently matched with TANF data to view the work status of those previously on TANF. An interviewee from one state mentioned they have a data sharing agreement to access UI data, but requesting the data is still difficult and it is not very accessible. The level of access depends on the language in the data sharing agreement. Another interviewee acknowledged they receive employment data to potentially match to their program data, but they do not reciprocate by sharing their program data. A handful of interviewees expressed minimal sharing of employment data across agencies within their state, noting legal issues that often require that programs' data remain within their own data system.

Many states include UI earnings and employment data in an integrated database, which presents a variety of options for using the data. As mentioned, some states use the integrated employment data for

⁶ The Work number database is a user-paid verification of employment database run by Equifax Inc. It is available at <https://www.theworknumber.com>.

⁷ The National Database of New Hires (NDNH) is available at <http://www.acf.hhs.gov/programs/css/resource/national-directory-of-new-hires-data-elements>. The website contains a description of the NDNH. A full guide to the NDHD is available at <http://www.acf.hhs.gov/programs/css/resource/a-guide-to-the-national-directory-of-new-hires>

verification of eligibility for other benefit programs. However, several interviewees noted they were able to perform more sophisticated analysis and data linking with this rich set of integrated data. The integrated data or data warehouses allowed many administrators to look at current and longitudinal earnings alongside receipt of TANF, SNAP, and sometimes Medicaid. One interviewee mentioned matching TANF and UI data once per year to see what happens to TANF leavers, which is a topic of great interest to many states and will be discussed in the section on research. One researcher also mentioned obtaining other sources of income from Census data, which can then be linked with TANF, SNAP, and employment data to get more complete information.

The use of UI earnings and employment data posed challenges for state and local administrators as well as researchers, challenges which were echoed during most interviews. The most common challenges included various limitations with UI earnings data, such as earnings being reported quarterly and detailed information about wages and benefits not being included. Additionally, there are certain populations for which earnings are not available through state UI data, including earnings out of state, earnings of military personnel, and earnings from employment that is not reported to UI. As a result, states understand the data they have may not be as complete as they would like. A local agency interviewee noted one other challenge with the state employment data: because of their state tax structure, the data they receive is only for their state-based employers (if corporate headquarters do not reside in-state, they do not have the data). One researcher stated that using IRS data for earnings is also not complete as it only contains what is legally required for employers to report and omits some income people receive, such as pre-tax benefits like health insurance and transportation. This researcher would like to link data in UI earnings records with tax data to get a sense of how much is missing.

As first discussed in this section, most states and localities have access to some level of UI earnings or other employment data. However, two interviewees indicated that they currently did not have access to this data but would like to have this data. Other states that have more limited access (like those detailed above) indicated they want more access or would like more information to be gleaned from the earnings and employment data they do have. Many of these interviewees want more detail about earnings (when employed, hours worked, fringe benefits, and out-of-state earnings information) as well as additional labor market information (what jobs are available, required skills for people to connect to jobs, and employers' data). As one interviewee stated, "we are getting savvier about getting UI wage data," having hired a full-time data person recently to help in data extraction and linking efforts. Additionally, several state administrators want earnings data from a variety of time points—real time earnings as well as earnings histories, along with earnings at program entry, during service, and upon program exit. Other examples of ways states and localities would like to utilize and advance their use of employment data

include: integrating WIC, DOL, and IRS data in the state's data system; matching UI earnings data with corrections data; and looking at earnings as part of family stability, such as the percentage of income used for housing.

Nearly all of the researchers and other data users with limited access to this data would also like to have more data available and do more with the data they do have available to them. Specifically, state administrators noted they want more detail about the jobs themselves and more information about earnings, such as how much time individuals spend working, benefits and non-reported income, and wages earned across state lines. Others want more widespread and consistent access to Department of Labor data. Two interviewees also discussed the barriers that researchers wanting to use NDNH data experience; for example, currently, if a project is not federally funded, researchers cannot access that data.

Medicaid and Health Data

Many of the administrators from states and localities that were interviewed mentioned using Medicaid and other health data to some extent, but only a few of the researchers specifically cited access and use of these data. For many of the state interviewees, Medicaid either falls within their agency or is part of the state's integrated benefit service (for example, if an individual is on TANF, they are automatically enrolled in Medicaid and SNAP). Similarly, a few interviewees acknowledged having access to the Medicaid data because Medicaid eligibility determination falls under their purview. For these interviewees, the data were easily accessible. Other interviewees had access because of partnerships with the health authority, the state department that administers Medicaid, or hospitals, which can aid data sharing agreements.

For a handful of states, these data are housed in the same database or data warehouse as other public benefit programs. As previously mentioned, these integrated client databases often include Medicaid in addition to SNAP, TANF, and SSI data in the state. However, a number of states specified that the Medicaid data is kept in a separate data system, even if the state has a data warehouse containing integrated data. One such state cited Health Insurance Portability and Accountability Act (HIPAA) restrictions as the reason they have a well-linked data warehouse with everything except Medicaid data. There are projects underway to integrate the Medicaid (and Children's Health Insurance Program, or CHIP) data with other agency programs in some of these states. For those with an integrated database, or automatic enrollment based on receipt of TANF, linkage between the systems' data is inherent in the data system. For other states, linking data with Medicaid and health data can be more challenging, but ad hoc linking with health services and Medicaid data is occasionally performed.

The handful of researchers that have access to Medicaid data mentioned using it to look at movement in and out of programs and to study health outcomes for children in households receiving TANF. The

researchers that have access to Medicaid data tended to have generous levels of access to numerous state administrative datasets, usually due to their organization or university's relationship with the state agencies. Thus, these researchers have access to Medicaid eligibility and claims data when needed for their work.

There were only a handful of references to using other health systems' data, including substance abuse and mental health data. One interviewee said they use a centralized assessment screening—which includes mental health, substance abuse, and other health issues—as part of their intake process. Another interviewee noted they had access to all state administrative data from mental health and substance use services, as well as the state employee health plan system.

Several states expressed interest in having access to health data, or access to more health data, to improve their work. Data concerning substance use services received, health indicators, and community health rankings would provide a more complete picture of the family's level of sufficiency and use of services. “In order to look at the family as a whole, for our family-centered service approach, we need access to all the data [for the services] that families are engaged in,” explained one interviewee. Another interviewee noted they want to receive CHIP data and others mentioned a desire to better integrate the Medicaid data into their existing database, in order to access the health and Medicaid data that is now currently separated from their programs. One interviewee mentioned that if the systems were integrated, they could document the presence of physical, mental health, and substance abuse issues at entry, during service, and upon program exit.

Additionally, a few state administrators noted that while they have access to Medicaid data for eligibility purposes, no other use of the data is possible. One interviewee, stating her agency does not have access to Department of Health data clarified, “We're required federally for programs to have an income eligibility verification system. So when I mean [I want health] data, I mean for research and outcomes analysis. I can check one time if [a] child has been immunized, for example, but I can't get the data at the aggregate batch level to do the research I need.” Along similar lines, another state administrator mentioned they would like access to the Federal Data Services Hub (the Centers for Medicare and Medicaid Services (CMS) database used for determining program eligibility) for use in research. However, the Hub is currently only available to states for eligibility purposes. Finally, one state interviewee acknowledged that while they have a good relationship with the department that oversees Medicaid and understand the potential uses of their data, they currently use Medicaid data only to determine eligibility and not for broader purposes. “Our commissioner would like to use Medicaid data and match with them to see who is using Medicaid and not TANF and who would be eligible. But it's not high on her priority list right now,” perhaps due to time or resources.

Education and Training Data

Education data are used very rarely by the individuals in our interview sample. About one-fifth of administrators interviewed mentioned using education data, but in almost all cases the data used is in aggregated form, or there is a data sharing agreement with the Department of Education that is not frequently used. Very few administrators mentioned more substantial data sharing with Department of Education data and only a few researchers noted they used education data.

Despite the limited access to and use of education data, a couple of interviewees did speak about how they access or use education data. One administrator uses school attendance data to see if children more regularly attend school as their families become self-sufficient. A local administrator explained that they can access data from the county's largest public school district but not from other districts. In at least one state, the Department of Education maintains a data warehouse containing K–12 and postsecondary student data integrated with public assistance data, incarceration, earnings, and assets. These kinds of warehouses may become more common in the coming years as a result of the Department of Education's Statewide Longitudinal Data Systems Grant Program,⁸ although no interviewees specifically referenced the program.

However, education data is among the data types that interviewees most often noted they wanted to access. More than half of administrators and about one-fourth of researchers wanted to use education data. Usually, respondents said they wanted to access education data in order to understand the educational histories of their adult clients or the current experiences of their clients' children.

Education data can be challenging to access and use. Several researchers discussed their experiences trying and failing to access these data in the past, because of either legal restrictions or the perception of legal restrictions to sharing the data. A couple of interviewees also noted that because education databases tend to rely on their own unique identifiers (rather than SSN), this data can be more challenging to link.

Interviewees expressed contrasting opinions about whether human service programs should in turn share data with education departments. Some thought that it was important for educators to know which families are at high risk, since children's education is a key element of family well-being. Others were concerned that a child could be stigmatized as coming from a family receiving public assistance.

In addition to education data, a few interviewees specifically mentioned using job training or other workforce data. Some administrators also wanted more workforce data, but their focus was generally on better understanding the outcomes of workforce program participants.

⁸ For information about this program, see <http://nces.ed.gov/programs/slds/>

Child Care Subsidy and Child Support

About one-third of the state and local interviewees mentioned having access to and using child care subsidy data, while only a few researchers and others interviewed have access to such data for their work.

Slightly higher numbers of state administrators reported using child support data, but only one researcher acknowledged use of these data. Administrators use these data for a variety of reasons, often because they are responsible for child support enforcement in their state. In one state, the agency looked at how many individuals receive child support through their administrative records. It then identified categories of recipients, such as those that cycled on and off or those only getting partial support, and whether they could engage the cases with large arrears. A researcher who uses child support data had access to a rich set of administrative data in the state and noted their work is around implementation of how programs work together, including child care and child support.

Of the interviewees that have access to either child care subsidy or child support data, several noted that the data are kept separate from their agency or department's main database. Child care subsidy data is often detached from other programs, sometimes being the one set of data not included in a comprehensive integrated database of administrative data for the state. There were a few examples of states that keep their child support and child care subsidy data in an integrated data warehouse, however, and others are actively trying to do so. One state interviewee mentioned that while there is currently not a data warehouse, the Departments of Medicaid and Human Resources, along with the Governor's Office of Information Technology, are working together to integrate TANF, SNAP, child care, Medicaid, and CHIP.

Because child care subsidy and child support data are often kept separate from the majority of other public benefit programs, linking and data sharing with other agencies' and programs' data is infrequent. There were several cases of states noting that sharing of child support data, in particular, is extremely limited or nonexistent. Some indicated that in their state child support data could not be shared legally. Some interviewees mentioned having access to child support data with the proper data share agreements in place, but even with the agreements the data they can access is often limited. One interviewee felt that as a result of the limited sharing of child support data with other human service agencies, child support is "not really considered in discussions in [the] overall picture of human services for [the] poor. [There is] not a lot of meeting of minds on coordinating [with other agencies]." Additionally, because the population moves around frequently, there are many different data entry sources into these systems, which can lead to errors: "bad data overrides bad data" in the child support data system, according to one respondent.

A few interviewees mentioned wanting child care subsidy and child support data. Some do not have access to the data at all and others who use it already would like more integrated data. One admitted

wanting to use the child care subsidy data to better understand the impact on different families of receiving the subsidy. One state administrator mentioned that the child care subsidy data available to them is aggregated and not available at the individual level, as they would prefer. Additionally, the subsidy program is split between two departments. Due to a lack of cooperation between the two, there are problems with accessing child care data. Another interviewee wanted to access child support data from other states. A third respondent noted that his agency can only access the child support data for families currently engaged with their programs, but access to the full dataset would be beneficial.

Child Welfare

Over half of the states and localities interviewed, and a few researchers, discussed having access to child welfare data in their state. For example, one state described the development of a performance management tool that uses eligibility systems and child welfare data to do data mining and examine how well programs are operating across the state.

While some states have the child welfare data housed in a data warehouse with other programs, the majority do not. Like child care subsidies and child support data, child welfare data is often maintained in a separate data system from the rest of the human service programs. One interviewee described their state's human services as being split between the income maintenance and social work worlds (where child welfare and foster care reside). The state keeps the data from those two worlds in different systems.

Despite having separate systems, there are several examples of child welfare data being linked with other administrative data, specifically SSI, health services, education, the homeless management system, and other interagency programs. One interviewee discussed an existing link between the Departments of Human Services, Child Welfare, and Housing. This link allows the departments to focus on identifying families they have in common and to jointly create strategies to serve this population most effectively.

Another state discussed a program within the Department of Workforce Services through which they can receive social security and child welfare data. Another state is currently creating an Early Childhood Longitudinal Data System to integrate data across child welfare, health services, and education, allowing them to see where the three systems have overlaps. In creating this system, the state is in the process of creating a unique ID to be used across all program areas, so they no longer have to manually link datasets, which will save significant amounts of work. This state, while acknowledging child welfare has “the most resources,” also admitted that historical child welfare information is not conveniently warehoused.

However, state workers indicated that the data is stored and not easily accessible, requiring a request to the Department's IT team. Overall, the sharing and accessing of child welfare data often requires a data sharing agreement between departments.

The biggest challenge for many interviewees working with child welfare data is linking to other systems, because, in the interest of protecting the children they serve, these agencies often do not use social security numbers. Not all interviewees struggled with linking, however. One interviewee said her biggest concern is with the quality of the data in the child welfare system. Due to the high volume of work that needs to be done with caseloads, case managers are unable to focus on data entry. Thus, the data entry becomes an afterthought. The interviewee explained she always warns others using the data to “take it with a grain of salt,” as trained social workers don’t have time to prioritize data entry.

Several interviewees discussed the child welfare data they would like to access. One would like to see more connections between the education and child welfare systems. Another indicated that even within their own department the data is often not shared; foster care, TANF, and Medicaid are legally restricted from fully sharing their data in the state, and it can only be shared on a case-by-case basis. One interviewee that does have access to these data mentioned the backlog of initial assessments in child welfare and their desire to study the impact of this backlog on children. However, they do not have access to data at the individual level, which makes determining the impact of the backlog difficult. A number of states and local administrators mentioned existing projects underway to create a database system that will include child welfare data, some of which were discussed in prior sections of this report.

Criminal and Juvenile Justice Data

Administrators and researchers use criminal justice data to better understand the populations they serve, but they do not use this data as commonly as other data sources. About one-third of administrators and a few researchers mentioned using criminal justice data. The type of criminal and juvenile justice data most frequently used are incarceration records integrated into broader human service data systems. A few interviewees mentioned using arrest records or data on court charges and convictions.

Interviewees did not speak about the specifics and challenges of using or linking criminal justice data. Because we did not interview any administrators working in criminal justice, we also did not hear about experiences sharing these data. A handful of administrators and researchers do not currently use criminal justice data but would like access to these data. Court records can be particularly helpful in identifying child-parent links and family connections across households. However, interviewees indicated they mostly want to see if clients had an incarceration history or if former participants were incarcerated.

Earned Income Tax Credit and Tax Data

About a quarter of interviewees specifically mentioned wanting to know who claims EITC, and several interviewees wanted other tax data to help them get the best possible understanding of families’ financial circumstances. However, only two respondents had actually used tax data. One of those respondents

indicated only individuals working for or with the IRS or the Census Bureau can access tax data. All interviewees who mentioned tax data agreed that access to the data is the biggest challenge.

A couple of interviewees mentioned they had worked with tax data in aggregated form. For example, one administrator had worked with his state's Department of Revenue to measure EITC take-up rates using metrics on EITC applications for particular populations, such as SNAP participants or participants on both SNAP and TANF. However, this administrator was able to see the numbers only in aggregated form and did not have access to whether specific individuals had claimed the credit.

Census and National Survey Data

About one-third of respondents mentioned using public data from the Census Bureau or other large national surveys. Specific surveys mentioned include the Current Population Survey (CPS), the American Community Survey (ACS), the Survey of Income and Program Participation (SIPP), and the National Longitudinal Survey of Youth (NLSY). In most cases these data are used to understand the demographics of poverty in a particular area. For example, one interviewee mentioned that her department measures their performance at scale using poverty measures for the region. Another administrator uses the poverty rates across communities to determine where service utilization is higher or lower than expected in a given area.

A handful of interviewees want access to survey data linked to administrative data, both to provide additional context for administrative data and to assess the extent of underreporting for benefits receipt in the surveys. Respondents identified this underreporting as a particular challenge in using survey data for FSS work.

Original Data Collection

Many of the administrators interviewed indicated they utilize original data collection in their work for a variety of reasons. Specifically, they noted states implement surveys and original data collection efforts to supplement their administrative data and meet their research and performance management needs. The decision to collect data was frequently motivated by a desire to understand client outcomes in areas where administrative data are not available. Four interviewees mentioned surveying TANF leavers to get a better understanding of clients' self-sufficiency and potential engagement with other human service programs after leaving TANF. One state used this data primarily to look at wages 13 months after program exit, but also used it to compare recidivists and "permanent" leavers. This interviewee also conducted a survey of the TANF returners' EITC knowledge. A few other interviewees also mentioned their use of survey data and focus groups to get data on outcomes, primarily at the client level.

States also conducted surveys to meet other unique data needs that couldn't be met by other sources. Interviewees' examples of this kind of research include surveys to examine the financial goals and behaviors of clients, a survey for a cost of living study, and a survey of child care providers to discern some of the out-of-pocket expenses that may make child care cost prohibitive. One administrator said that when it was required his agency had surveyed child care providers to assess the market rates of child care. Finally, states collected original data in order to improve program efficiencies or to learn from other states. One interviewee explained they engaged in original data collection via Internet and phone research in order to learn more about what was going on in other states and to find out how and why, in certain areas, states were successful in moving families towards self-sufficiency.

Several researchers also relied on survey and other qualitative data to supplement the quantitative administrative data. Many of the researchers talked about their ability to link their original data collection with administrative data to strengthen their work and allow for comparison, outcome, and impact analyses that would not otherwise be possible. Examples of the studies researchers pursued with surveys and evaluations were varied, but included the following: employment outcomes of nonworkforce programs, the difference between food stamp records and who actually lives in the recipient's house and consumes food, linking survey data to child welfare nonprofit contracts with the state in order to look into how fiscal standing is related to organizational management, and why people apply for assistance but do not enroll in a program.

Interviewees generally did not discuss sharing survey and other qualitative data. However, one interviewee raised the idea of exploring ways to use survey data for multiple projects. Given that survey data can be rich with information that cannot be obtained from other sources, it could be useful for multiple projects. However, being able to use the data might be difficult because of security considerations, IRB restrictions, and budget limitations.

The primary quality concerns with original survey and qualitative data are low response rates and underreporting of earnings and benefits receipt in surveys relative to administrative data. Interviewees gave a handful of examples where this is particularly true, such as high nonresponse rates for questions about income and assets or program participation. There are generally discrepancies in earnings data obtained from surveys when that data is compared to administrative data earnings estimates. A few interviewees indicated they do not use survey data due to concerns about the quality of the data. These respondents warned that administrative data is much more accurate than survey data, with one stating that administrative data are "the real facts."

While many states administer their own surveys, several state administrators indicated they also would like to use more survey research to fill gaps in their data and research knowledge. Some expressed interest in survey data that can look at the impact of a program and whether the program has added value. A few interviewees commented that surveys are needed to get a more complete picture of an individual's well-being and to get at some answers that administrative data simply do not measure. For example, one interviewee said, "There's a gap [in knowledge of what is] a subsistent budget where everything is paid for by [the] family. Knowing what income is needed, or public benefits needed, is a good start. And [we] can't know that all by administrative data, we . . . need survey data." Other examples of where surveys would add value to work include surveys of parents on sources of economic insecurity as well as character studies of the individuals enrolling for benefits.

Interviewees discussed the expense and complexity of original data collection. Specifically, there is a need for more staff and financial resources to create, administer, and analyze survey or evaluation data. This is necessary whether the work is done in-house or contracted to a third party such as a local university. As previously mentioned, good data costs money and collecting data takes a significant percentage of the budget—funding that is often diverted from direct service.

Other Data

Interviewees used data from a variety of other sources, particularly administrative data from other human service programs. These programs included state-funded disability, energy assistance, substance abuse prevention or treatment, data on public housing and homelessness programs, and administrative data for specialized assistance programs targeting veterans, refugees, the aging, or immigrants. A few respondents also mentioned using consumer expenditure or market research data.

Other datasets respondents wanted included data on particular populations, cost of living data, other financial data, and birth records. Interviewees wanted better access to data on veterans, the homeless, and the disabled; several mentioned access to SSDI data in particular. In order to understand a family's cost of living, they wanted to better understand housing and transportation costs. Other data requested include credit scores, insurance or risk assessment data, and consumption data.

In general, interviewees also wanted access to data that they could use to better understand their clients' circumstances, including data about demographics and barriers on a range of topics, including income, nutrition, family situation, domestic violence, and substance abuse. Although these other data sources were noted in our interviews, they were mentioned by less than one-fifth of those interviewed.

Data Quality

Regardless of the types of data states, localities, and researchers used or emphasized that they wanted to use in their work, there are some crosscutting concerns regarding data quality that emerged in our needs assessment discussions. In addition to the challenges with identifiers and linking data mentioned previously, interviewees had concerns about data entry errors and inconsistencies, incomplete data, understanding the context of data collection, and data timeliness. Interviewees offered numerous suggestions of ways to overcome these challenges. There were also a few interviewees who claimed that some or all of the data they worked with was of very good quality.

Data Entry and Accuracy Concerns

Most prominent among the quality concerns were those related to accuracy of the data and data entry errors. About one-third of those interviewed expressed concern about data for a variety of accuracy and data entry reasons. For starters, because most data systems discussed by interviewees are eligibility systems and require manual data entry, they noted that priority is placed on the data fields relevant to eligibility. Data in other fields are often not validated and are frequently missing, incomplete, or incorrect. According to several interviewees, if a data point is not essential in the provision of benefits, the more questionable it becomes. The example given many times by interviewees was the quality of client educational attainment data. While many systems have a field for this information, the data are often not collected, entered incorrectly in order to move to the next data entry screen, or not updated after the intake or application process. “Where data entry isn’t required [of the] caseworker, the quality is not as good,” one interviewee summarized. Additionally, many interviewees noted that because caseworkers are focused on service and face increasing caseloads, data quality and the timeliness of data entry often take a back seat to other tasks. Instead, caseworkers enter the minimal required data needed to enroll clients. As caseloads are increasing, frontline staff prioritize service and benefit allocation at the expense of data quality.

Respondents also noted that inconsistencies arise when there are hundreds or thousands of different people entering data into the same system. Often, there are differences in how staff enter data or interpret data fields. Interestingly, all three of the interviewees who made this point come from agencies in states where the system is state-supervised but county-administered. These systems likely see more variety at the county or local area level in how programs operate and enter data into systems. Further complicating data quality, agencies often use multiple systems to obtain the data they need. If data is entered differently in multiple data systems, errors occur that require staff time and resources to address.

Finally, many states and counties are working with old data systems that need modernizing. This can also contribute to data entry errors. Older systems require more manual data entry, reentry, and manipulation, creating more opportunities to introduce error. Paper intake forms, Excel spreadsheets that must be converted to report form, and certain data elements that must be reprogrammed or recoded were all examples interviewees gave of error-prone processes.

Incomplete Data

Another common quality concern expressed by interviewees focused on the prevalence and challenges of data that cannot be comprehensively collected. These included, but were not limited to, datasets missing certain populations, data only available in aggregates when individual-level data is needed, and the absence of historical data. Underreporting and nonresponse in surveys, discussed previously in this report, are also missing data concerns noted by interviewees. A few interviewees referred to incomplete data from individuals who move in and out of a state. Historical data is missing for these individuals in their new location, and in the states they leave. This movement introduces ambiguity into treatment estimates that aim to adjust for program attrition: did the individual leave the program, or did they continue to receive benefits in another state?

Understanding the Context of Data Collection

Another common quality issue raised was the need to thoroughly understand data sources. Several interviewees emphasized the need to know how data are entered, look deeper into exactly what is collected, and understand the origins of what the data mean. “You have to go into the agencies and understand how people are entering things,” one explained, in order to understand why data may be missing, incorrect, or duplicated. Because data field definitions are constantly shifting and can be different from one agency or provider to the next, understanding the definitions is critical. For example, SNAP has rules about what they do and do not count when collecting income data, so they may capture different income information than other data sets also collecting data on income. Detailed documentation is needed to answer questions such as why the data may be missing, what fields are edited, and how to handle duplicates. Interviewees also emphasized the need to have background program information from

someone who knows both the program and the data in order to become aware of and overcome some data quality problems. This knowledge tends to not be well documented.

Timeliness

A final quality concern that was mentioned by the interviewees was the issue of timeliness in receiving data and inherent lags in data. Whenever possible, administrators noted that states try to create real-time access to data. However, some states are better able to do this than others. The states and departments that have slower, older systems and data entry processes tend to have more data delays. Additionally, certain departments are notorious for data lags for at least 3 months or more, for verification and other reasons. UI, corrections, and child support data were all repeatedly mentioned as having longer lagged data. When data are not available in a timely fashion, it impacts the type of analysis as well as the accuracy of the data available, as data such as income may be outdated by the time it is ready for use and analysis. Household income and expenses, for example, are often variable throughout the year, so capturing annualized data misses seasonal nuances. During the last economic crisis, when conditions often changed quickly with a family's well-being and their ability to be self-sufficient, waiting for data created an issue for many states, localities, and researchers trying to use data to help inform effective practices.

Addressing Quality Concerns

Although interviewees raised several data quality concerns noted above, many had feasible and innovative suggestions for addressing the issues. Many states noted it would be beneficial to train staff to better understand the value of data and the process of data entry. Part of that training should involve explaining specific data fields. As one interviewee noted, "We do better when caseworkers know the factors that go into reporting, because then they do better at entering those [data]." This training should be directed towards caseworkers and those entering data, but one state indicated starting with midlevel managers to impart the importance of data quality, assuming that emphasis would then be conveyed to frontline staff. Many others offered suggestions to improve the tools caseworkers and data entry staff have to execute their work. The most common example is improving the system in which they enter data, which will make data entry more accurate and efficient and help reduce errors. Additionally, administrators from two states noted they were considering giving caseworkers a portable tablet to do real time intake and assessments, allowing for more timely data entry and reduced data entry errors.

According to interviewees, many states and local agencies also undertake routine quality checks and data cleaning efforts. The most common anomalies and omitted data occur with the identifier data field, which these need to be cleaned before running analysis. One state gets a list every quarter of all inaccurate social security numbers (SSNs) in their data. Another interviewee mentioned his team ensures they do

continuous maintenance with the identifier fields, ensuring everyone has an SSN or, when an SSN is unavailable, a pseudo ID. In addition, another interviewee mentioned that part of the process of moving their data into an archive housed by an outside vendor is that the vendor uses a data cleaning tool to look for frequent misspellings, consolidate values, and look for other errors. While many states have quality checks in place, one administrator mentioned that for the most important data fields, staff are assessed by how accurately they are entering data.

Despite all of the data quality concerns expressed throughout the interviews, there were several interviewees who pointed out that either some or all of their data is actually of high quality. One interviewee responded he had no quality concerns with his former agency's data, saying that "it was perfect. It was our data; it was from administrative records that generated the benefit so it was high quality." Several interviewees discussed how because eligibility and payment systems have many quality checks to avoid accidental mistakes that could prevent or delay an eligible client from receiving benefits, the data are generally quite good. As mentioned previously, the more a particular data element is used in administering and granting eligibility for a program, the more accurate it is. Aside from the eligibility systems, some interviewees noted other systems with higher quality data, such as wage data which is verified each month. Another interviewee indicated the child support data relies on automation more than other data and is therefore more accurate. Overall, some systems are more reliable than others, and this can vary at the county, agency, or even program level.

Research and Data Needs

Research Questions

Throughout the conversations during the needs assessment, the states, localities, and FSS researchers discussed an abundance of research questions they would like to be able to address and study. These can be grouped into 10 broad questions, which are listed below along with a few examples of the research questions or topics that were mentioned for each category.

How effective is this program and what are some impacts and outcomes on clients?

- What impact is our program having on FSS and well-being? Are people better off, are we pulling them out of poverty? Has our program added value? If not, are there aspects of the program that work better than others?
- What are some interim outcomes we can look at to monitor clients' progress and make process adjustments?
- How do performance management policies impact street-level caseworker practice and interactions with clients?
- Are programs working as designed?
- What are the long term savings to the system from a family or person's engagement in a program?

How does receipt of multiple public benefits impact a family, and what is the composition and impact of those multiple programs?

- What other services are families receiving? How do all of the programs interact together, and what is the impact on families receiving these other services? The interaction of TANF and SNAP recipients with EITC, child care subsidy, child support, transportation and housing assistance were referenced the most.

- Is there a right combination of services (“a success package”)? What works for different families? Does any combination predict success for families with certain barriers and characteristics? Who needs which services?

What are the outcomes of individuals that have left TANF?

- Are families that have left the TANF system better off?
- What other services do TANF leavers utilize?
- What are job growth patterns and earnings of program leavers over the long term? What are the postemployment services that prevent returns to TANF?
- What has happened to the leavers who “timed out” of the system? How does their trajectory compare to those who left because they got a job?
- What characterizes the individuals who return to TANF?

What are the outcomes for special populations engaged (or not) in various FSS programs?

- What do outcomes look like for: the disabled? Veterans?
- What are long term outcomes like for children who grew up on TANF?
- What are outcomes like for youth aging out with their own children?
- What happens to families in poverty that don’t receive assistance or don’t apply for programs? What about those who applied and were turned down?

What are the characteristics of the population in specific programs and what motivates and impedes their success?

- What does our population look like (by geography and type)?
- What barriers do program participants face?
- How does the current pool of program participants differ from previous groups, and how do services and outcomes vary with those differences?
- Generally, what motivates people to succeed?

What are the impacts of economic and policy changes?

- What is the effect of the minimum wage on poverty?
- How do workforce development programs impact poverty?
- How do poverty counts vary before and after taxes?
- How do changes in the national economy impact TANF demography?
- What is the value of the investment in the local economy? Is it enough to justify increased expenditures?

What is a living wage and what is included in living self-sufficiently?

- What's the "livable wage"? What is the cost of living in a particular area? What income is needed to be truly self-sufficient, including the costs of transportation, healthcare, education, and other necessities?
- How does cash flow impact clients?
- What is needed for long term self-sufficiency? Who is succeeding in achieving long term self-sufficiency?

What are the differences among the various groups of participants?

- Who are the different groups of participants, and how are these differences important for understanding how to engage them in programs and what support services are needed?
- Where there are small groups of families that use a high proportion of the resources, who are they and how can resources be used to address their needs holistically?

What employment supports and benefits most commonly lead to retention and achieving self-sufficiency?

- What are the long term impacts of employment benefits, part-time vs full-time work, or the number of hours worked by a program participant?
- Do work supports encourage or disincentivize work?
- How do you measure success in work participation?
- What jobs are available—now and over the next decade—and what skills are needed for those jobs?

How does my program compare to its equivalent in other states?

- How do outcomes compare across states? What variables contribute most to differences in success?
- What do other states look at in the way of outcome measures? What general trends are they seeing in certain program areas?

Data and Tools Needed to Address Research Questions

Data Needed

Throughout these discussions of what research the interviewees would like to address, there was always a conversation about the data that would be needed in order to pursue these research questions. Below is a list of the most commonly referenced data that states and researchers would need in order to conduct their research:

- Data on participants once they leave the system
- Longitudinal data
- Data across programs, such as integrated data from all services and systems with which program recipients engage

- Data across states, for comparing their program data to other states and tracking families that move across state lines
- Client survey data, to obtain difficult to measure factors like motivation
- Data to construct appropriate comparison groups
- Measures of family well-being
- Organizational and management data

Tools Needed

In order to answer the research questions identified here with the data they would like to use, additional research tools would be needed. Interviewees suggested tools including a data warehouse, which, as previously mentioned, many states already operate. Other tools include better and more uniform assessments throughout the state, a public-facing database, and resources such as a call center and mobile devices for caseworkers to improve and streamline data entry. Other research and analytic tools that would aid the research include improved sharing of data within and across states, tools to more effectively present research and information, reliable cost-benefit analysis tools, and analytic tools to better identify people with whom to conduct program outreach.

Feedback for the FSS Data Center

Better Use of Data in FSS

Interviewees were asked general questions about how data could be better used in FSS policy, research, and administration. Their answers reinforce the value of the Data Center’s mission. On the whole, respondents stressed the importance of increasing the accessibility and usability of data. They also stressed the importance of improved communication between administrators, researchers, and policymakers. These are important key goals for the Data Center.

Overall, interviewees agreed that there needs to be better education about data, including how it can and should be used. Data transparency was often mentioned as being critical to enabling government, researchers, and agency staff to better use data. One interviewee remarked that to better use data, the data users—agency administrators or staff, policy directors, governors, and others—must be aware of the data that are available to them. Of course, this improved education of the general public and transparency of data requires enhanced communication with data users. Having common definitions across programs and data using groups will ease this communication.

Interviewees also noted that data needs to be more available for public use and presented in a way that makes it usable for the general public. This includes educating the general public about the empirical knowledge that is available and explaining how it impacts them. One interviewee stated, “One thing to have around data is transparency. . . public availability of data is crucial, it does really good things.” If agencies better understand the data that is available to them, it may open the doors for greater transparency.

Many interviewees felt there were several ways data could be better used by government leaders, mostly to inform policy and improve decision making. In order to do so, government leaders need context around their data. For example, TANF helps families move toward self-sufficiency, but TANF on its own will

not lift a family out of poverty. Throughout the interviews, it was suggested that government leaders must create a data culture to build data capacity in public agencies, which would allow data to be used to improve performance management and decision making. Interviewees emphasized that data should be used by government leaders to measure performance and track outcomes, to justify decisions and policies, to inform policy change, and to implement new programs. As one interviewee simply stated, “[For] any decision that can be better informed by data, data should be used.”

Interviewees felt researchers could better use data primarily to better inform practice. A few echoed the need to have a productive partnership with practitioners to do this. Dialogue between researchers and practitioners would help to ensure the research is useful to agencies and inform practice. Interviewees also felt the use of mixed methods of both quantitative and qualitative analysis should be encouraged among researchers, as should a focus on applying longitudinal methods to their data analysis. Interviewees suggested that researchers should also allow for more analysis of trends and different programs’ effect on families and that this can be done by developing indicators for program improvement. Finally, it was suggested that researchers interpret the data in ways that are meaningful for the government, so that results are presented in understandable terms that get the point across. It “all has to do with translation,” according to an interviewee.

Lastly, interviewees felt frontline staff working with the families could use research and data to better identify families most in need, help them ask the right questions, and identify people for certain outreach and services. Interviewees also noted that data can be used to identify trends with clients, allowing them to act proactively to help prevent relapses or inefficient use of services. A common suggestion was to keep the client in mind, since all data users want to use the data in the best way possible to ensure programs are helping the families they serve most effectively.

Role for the Data Center

Nearly all interviewees agreed the FSS Data Center could play crucial roles in helping them achieve their research and data needs. The most common role for the Data Center noted by interviewees was helping the interviewees address some of their comparative research questions (as previously mentioned). They would like to look at outcome measures across other systems—whether in other states, at agencies within their state, or in localities in their state—which they hope will be available through the Data Center. The “Data Center would be a window into comparative research,” stated one interviewee.

Cross-state research was the most commonly mentioned benefit interviewees hope to acquire through use of the Data Center. Several interviewees mentioned wanting to track families that may have gotten services in other states and wanting to look at intergenerational poverty. Many are also interested in using

other states' information to benchmark their own programs, performing comparisons across states and analyzing general trends in program areas. To support comparative analysis and benchmarking capacities, interviewees noted that the Data Center should standardize differences in data across states to create consistent datasets. Doing this would make using data from different states much easier.

Interviewees also noted that sharing best practices with other states was an important need that the Data Center could help address. This could be done by learning what other programs have succeed, how other states are managing similar programs, how programs are measuring impact, and what other questions states are asking in their research. The Data Center could also serve as a community to establish contact with and ask questions of other states directly. Some felt that the Data Center would allow states to operate more strategically and could facilitate regional collaboratives regarding common policy issues and data access to explore these concerns.

Interviewees also saw a role for the Data Center helping to use, analyze, and translate their data more efficiently. Many noted needing this help because they do not have the internal resources—staff, expertise, time, or infrastructure—to perform all of these tasks on their own. Interviewees feel the Data Center can provide the tools needed for more sophisticated data trend analysis, analytics, and decision making support on the ground, which would also help translate data to inform practice. Additionally, many interviewees envision the Data Center assisting with improving longitudinal analysis as well as accessing more comprehensive linked data. Interviewees hoped the Data Center would be able to link administrative data with survey data, help with linking other raw data without personal identifiers, link data across systems within states, and link earnings data with program history data.

Another specific role for the Data Center could be to provide access to data they otherwise cannot easily access, from both other agencies and other states. Interviewees suggested that the Data Center could help agencies negotiate and reach data sharing agreements to promote data access. They also emphasized that the Data Center could help overcome many common challenges regarding confidentiality, brokering relationships with public agencies, and addressing aspects related to accessing and storing data.

Interviewees hoped that the Data Center could help “open walls” within and across state agencies as well as provide access to local data. Along these lines, data users interviewed hoped the Data Center could provide access to linked, individual-level data so data users could identify individual engagement across multiple programs, track individuals across assistance programs and over time, and obtain improved information on client demographics.

There were a handful of other suggestions that interviewees made for the Data Center, including:

- provide sample documentation and codebooks (interviewees often mentioned the need for a good data dictionary);
- offer transparency on where and how things function within a state, where to get data, and how data is collected;
- line up policies to see how programs are comparable, since understanding the policy is often required to understand the data;
- reduce duplication of efforts for states as they determine eligibility and enroll families and individuals in programs across different agencies; and
- serve as the warehouse to house the data ready for access and to save data from one project so it can be reanalyzed for other research.

Training and Technical Assistance

Most interviewees noted that a critical function of the Data Center will be to provide training and technical assistance (TA) to states and data users. Many interviewees also suggested the Data Center can encourage peer engagement among data users. Interviewees identified various training opportunities, acknowledging that the training topics should vary depending on the needs of the various types of data users. However, some clear suggestions for training and TA topics emerged from the conversations with all interviewees. These include technical and analytic training on a variety of topics, including statistical methodology, measurement issues, linking data, program evaluation, research methods, and how to manage and interpret data. Several interviewees referenced Chapin Hall's Advanced Analytics for Child Welfare Administration training as a good model for this type of training. Interviewees also noted that tying analytics and data to policy and practice would improve the ability of all staff to understand the data and its impact. Additional training and TA topics suggested included how to compare data across states, addressing data quality concerns, and how to offer more personalized TA to states to answer specific questions or topics of interest. Many interviewees mentioned that basic training on working with the data tools themselves is needed, including training on data use, access, what is available, what it can and cannot do, and how to interpret the data. Also, TA is desired on the practical use of data, such as how the Data Center can assist with states' reporting requirements and how the Data Center can support users in analyzing the data for its intended purposes.

In terms of who needs training and TA, there were a wide variety of responses from the interviewees. Overall, the vast majority noted they could use TA and training on the Data Center. Some felt certain types of data users should be trained before others, such as training administrators first, then case managers. Others felt the practitioners, case managers and those entering the data should be the primary audience for training. What is most important is that the training and TA be targeted and personalized for

the intended audience, helping to meet their needs and meet them at their level of understanding. While many interviewees noted that in-person training is ideal, the reality is that web-based training is likely a more common and cost-effective way to deliver training. Suggestions included starting with some in-person training to get states started, but then move to online or video training. They also suggested that webinars could include peer-led tutorials. Additionally, trainings could be clustered based on states' level of sophistication using data, as their needs and questions will be similar to their peers. One interviewee suggested taking advantage of national conferences to display information about the Data Center and hold in-person workshops and trainings onsite, which will reach large audiences at minimal cost. To augment training, interviewees also noted a toll-free phone line or call center for assistance with the Data Center would be helpful. In general, interviewees noted that all methods of training should be utilized to the extent possible in order to reach the maximum number of participants.

Many who discussed the topic of peer engagement agreed it would be helpful and useful for data users. They felt data users would find peer engagement to be beneficial for their work. Interviewees noted that peer engagement can happen on many levels—with other states, with other agencies within the state, and within their own agencies. They noted that peer engagement among frontline services providers is also helpful. Additionally, many interviewees felt peer engagement between researchers and other data users would be important—particularly between researchers and practitioners, data programmers, and administrators. Interviewees emphasized numerous reasons why peer engagement is helpful, including a desire to share best practices and the transfer of knowledge within an agency. Others noted that even with different levels of sophistication, all agencies share common challenges and peer discussion could help states more efficiently address these challenges. With FSS data specifically, the FSS field is expansive and involves many agencies, programs, and professionals. In the FSS field, as one interviewee noted, “each of us brings a new and different lens, different questions, [and] different ways to describe the work we do around family self-sufficiency.” Peer engagement would allow states to share strategies and provide examples of creative uses of data. As with training, peer engagement can also assist with technical challenges and help data users help each other navigate issues regarding data confidentiality and sharing. Finally, interviewees noted that peer engagement can also connect researchers with public agencies who need help analyzing data and understanding how research can inform practice.

Implementation Challenges

Although the Data Center was almost universally seen as a helpful potential resource for accessing, using, translating, and analyzing data, many interviewees cautioned us about challenges they anticipated in the implementation of the center. These largely revolved around getting states to participate in the Data Center—mainly for legal, security, and political reasons—as well as challenges comparing data across

states. On occasion, an interviewee did not see much need to use the Data Center to enhance their work because they already have a high capacity to work with data.

A handful of interviewees noted concerns states and data owners may have around releasing individual-level data. Some administrators, speaking about their own data, noted that there would need to be a discussion about confidentiality challenges and how to overcome them; at this time they did not think they would be able to share their data with the Data Center. They noted legal restrictions are in place that would limit sharing and linking some or all data. It was not that the technical aspects were viewed as problematic, as several noted, but rather people's apprehension about sharing data and allowing linking. These tasks will be politically challenging. With time and resource constraints, it may be difficult to get states to commit to engaging in the Data Center. Two interviewees noted that perhaps once there is a critical mass of Data Center participations, it will be easier to get states to commit; in the meantime, getting to that critical mass will be difficult.

Interviewees emphasized that the Data Center staff would need to serve as the custodian of the relationships and do the majority of the work to get states' data input and ready in the Data Center. To address this, several interviewees noted the importance of relationship building in order to obtain the data and facilitate the relationship with states. Two interviewees specifically mentioned that the history of Chapin Hall's strong relationships with agencies and their data, and knowing there will not be misuse of data, will help in establishing the Data Center.

Once data is stored in the Data Center, interviewees noted there are other challenges to consider. One includes addressing concerns over ensuring the accuracy and reliability of the data, as well as how to compare datasets that are not updated on the same cycle or using the same data elements.

Lastly, a handful of interviewees (4 administrators and 1 researcher) acknowledged they may not have much of a need for the Data Center, due to their current capacity and data availability. Though most acknowledge that they do need more information, or that there may be an area or two in which the Data Center could help meet their data needs, these interviewees felt their internal system largely met their data needs. Therefore, they would not devote resources to the Data Center.

Other Aspects

In addition to the roles, training needs, and implementation challenges for the Data Center, interviewees discussed a variety of other aspects for the Data Center team to consider as the project moves forward. First, several encouraged the Data Center to provide contextual background on the data—including information about how the data is collected and at what level—and be transparent about the data to users. Additionally, interviewees noted the importance of keeping in mind the many needs of the states as the

Data Center is developed and implemented. The Data Center team should think about the separate target audiences of data users, including who is invited to participate. It was suggested the Data Center team be very specific about how the Data Center can help all data users, emphasizing the infrastructure and expertise that will help users put together the datasets needed for their projects and research purposes. This may require developing a catalogue of common data fields across states and helping to understand what questions can be answered with the data. Specifically, interviewees noted that Data Center staff need to identify the outcomes and research states are interested in and which analytic tools they use, so as to be sure these are incorporated into the Data Center capabilities. Creating something easy to digest—like a dashboard—will help show states the practical uses of the Data Center and allow them to utilize the Data Center to its full potential. One interviewee recommended that the Data Center emphasize the capabilities for both practical applications of data for administrators and more general, exploratory analysis for researchers. Both functions are essential and needed in the final product.

Interviewees also offered their thoughts on how to market and share the Data Center, as well as potential competing or complementary sources of information to keep in mind during development. Several interviewees offered to help market and publicize the Data Center and resulting research. One interviewee suggested considering how the Administration for Children and Families' (ACF) Regional Offices can help disseminate information and engage users of the Data Center. Additionally, several interviewees mentioned existing projects and data warehouses that seem similar in scope and asked how the Data Center would be distinctive from these projects. One such existing project is the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan. One respondent suggested that the data should be fully integrated with new ACA data hub, which links health needs and mental health needs with FSS data.

Finally, there were specific questions around the Data Center that have yet to be answered, such as how much data will be included, whether there will be a fee for access or an expectation of membership support, and to what extent data cleaning will be part of the project.

Implications and Takeaways for Informing the Data Center

Translating Findings into Actionable Steps

The needs assessment findings will help guide the Data Center’s strategic planning, including building relationships with key stakeholders and designing and implementing the infrastructure of the center. In this section, we highlight key findings that will guide our work in the coming year, translate the findings into actionable steps for the technical development and practical uses of the FSS Data Center, and provide a guide for the technical assistance strategy.

In developing the technical assistance approach and the infrastructure for the FSS Data Center, we are guided by a number of organizing principles that emerged from the needs assessment interviews. First, it is important to address the needs and interests of multiple stakeholders and clearly identify how the FSS Data Center can be useful to each of them, while keeping in mind states’ needs. Second, marketing the Data Center and sharing ideas for collaboration with other entities is essential. Third, it is important to establish at the outset a plan for sustainability and how to support the FSS Data Center on an ongoing basis after development; the business model should be sustainable and fair. In addition, it is important to note that the FSS Data Center platform rests on proprietary systems, so it should be designed in a manner that allows the datasets and associated metadata archive to be transferred seamlessly and securely.

The needs assessment findings suggest a number of ways in which the FSS Data Center could support and facilitate family self-sufficiency research. Administrators discussed many areas that impact how data are used, including staff capacity to manage and manipulate data and the challenges of using for research systems that were designed for performance management. Respondents emphasized the importance of resources and relationships in using data for family self-sufficiency research, which the Data Center could augment for users. Other major concerns focused on the legal and security challenges associated with

access and data sharing. The FSS Data Center can provide tools to increase ease of use, accessibility, and the analytic quality of the data. The many tasks and roles of the Data Center can be organized in the following categories: data access and permissions; scope and types of FSS data; data security, storage, and documentation; data management, linking, and analysis; and training, technical assistance, and peer engagement.

Data Access and Permissions

There is clearly demand from many administrators and researchers for the Data Center to assist with the access challenges they face in acquiring the data they would like to use. These access challenges include legal and privacy barriers as well as limitations in the capacity to develop and maintain the needed data sharing agreements. The Data Center can help users overcome all of these barriers by performing the following functions:

- Assist with negotiating and establishing data sharing agreements (some respondents indicated this would be necessary in order to participate). The FSS Data Center may also be able to assist with executing data sharing agreements between states to respond to internal and external research questions.
- Provide information on state and local data infrastructure, enable access to data, and improve data transparency and clarity on what data is available from which sources.
- It may be necessary to purchase certain datasets if they are unavailable by other means. Given the demand for Medicaid and National Directory of New Hires (NDNH), a process whereby these datasets would be purchased and made available to states or researchers would have to be developed. Data sharing agreements between the original data provider, the Data Center, and the researchers would need to be created for this purpose.

Scope and Types of FSS Data

Data providers and users generally want to better understand clients' circumstances (such as demographics and barriers) and outcomes. To gain this understanding, they require use of multiple datasets to provide a comprehensive view of the clients being served. The FSS Data Center can assist by brokering access to datasets that will help provide this comprehensive view. The scope and types of some of these data include:

- Diverse data sets, including TANF, SNAP, Medicaid, child care subsidy and support, child welfare, criminal and juvenile justice, EITC and other tax, state disability, and SSDI and SSI data (the FSS Data Center would need to prioritize and stage these requests).
- Unemployment insurance and employment data, as well as more granular data within these datasets (e.g., earning, wage, hours, and benefits).

- Additional aspects of family self-sufficiency, such as adult and child education histories, child care subsidies and supports, and criminal and juvenile justice data.

The Data Center can provide a menu of the scope and types of FSS data that are available in FSS research. States and localities vary considerably with regard to the type and quantity of data that can be shared, whether internally or externally. A menu of the data that are accessible for research purposes will help all data users in their research and performance management efforts.

Data Security, Storage and Documentation

The Data Center will be able to securely store and document all data it receives, especially by performing the following specific tasks:

- Provide researchers with up-to-date documentation and codebooks and make them accessible to users. Needs assessment interview respondents identified this as a pressing need, which the FSS Data Center can address through organizing and indexing documentation and linking it to variables and record types in the data file.
- Facilitate secure storage, access, and use of data. Having the “right infrastructure and resources in place to appropriately store and use data” is an integral requirement of the Data Center to fill administrators’ data needs, as echoed by one interviewee.

Data Management, Linking, and Analysis

The Data Center can provide a range of services related to data intake, enhancement, linking, and dissemination. These services include ingesting new data; editing and creating needed metadata; indexing, archiving, curating and harmonizing data within and across datasets; deidentification of respondents and performing disclosure risk analysis; and providing a diverse set of analytical tools. More specifically, the Data Center can help users with these processes in several ways:

- Assist with linking and integrating data, as well as providing tools and technical assistance for linking data to other systems.
- Assist states and localities by sharing data and leveraging external researchers to increase capacity.
- Assist data users with data preparation and creation of modeling methods to understand the policy impacts from of program changes.
- Provide clear, linked, harmonized, and contextualized data suitable for longitudinal analysis and benchmarking. Data providers and users reported that TANF and SNAP were the most commonly used datasets for family self-sufficiency research, but that these datasets need to be linked over time within a given state’s database in order to conduct longitudinal research.

- Help states to link survey data they have collected with their administrative data. Some states/ and localities want to use survey data to fill in research gaps, especially since the data quality on key elements can exceed that of administrative data.
- Retrieve data out of legacy systems and make data available in common formats, facilitate data linking, and provide a robust high performance computing environment to conduct analyses.

Training and Technical Assistance

Findings from the needs assessment identified the topics, recipients, and modes of training and technical assistance delivery desired by potential users. Providing technical assistance is one of the core functions of the FSS Data Center. The purpose of the technical assistance (TA) and training is to use, analyze, and translate data more efficiently. It should be practical and utilization focused.

A number of potential TA topics were identified. We have categorized them keeping with the roles of the FSS Data Center team. The Data Center can provide TA around:

- data linkage
- trend, longitudinal, and comparative analysis (particularly to track or “follow families” over time)
- comparing data across states
- understanding the data itself, use and access of data, and its definitions
- data needed to answer specific research questions
- tying analytics and data to policy
- the Data Center’s practical use, including receiving, documenting, and addressing feedback from users in quality control efforts

The Data Center could also provide training and TA to help users navigate the many features of the Data Center. More specifically, they could provide assistance around:

- brokering the execution of Data Sharing Agreements
- preparing data for use, simplifying files, and editing for consistency
- technical and analytic tools
- the use of the system (access, manipulation, use of tools)
- confidentiality/statistical disclosure controls
- data aggregation

In addition to these topics, the needs assessment identified a range of potential stakeholders that would be served. The stakeholders included, but were not limited to, practitioners and administrators, state data analysts (to facilitate comparisons across states), case managers, policymakers, and sophisticated data analysts.

The needs assessment findings also indicate that TA should be offered to multiple stakeholders in multiple modalities in addition to addressing the range of substantive and technical issues described above. In keeping with ACF’s approach for the delivery of comprehensive TA, along with the range of TA modes suggested by stakeholders, we will provide TA at three different levels: universal, targeted, and intensive.

Universal forms of TA would be available to a broad range of stakeholders. They would include a toll-free phone line for assistance, an FAQ page, workshops about frequent challenges to working with state-specific or locally specific datasets, webinars, and displays, information, and workshops at national conferences.

Targeted forms of TA would be based on the common needs of multiple stakeholders. These forms of TA would include workshops for clusters of states based on level of sophistication around data use, online tutorials and videos posted on YouTube, and in person workshops at national conferences.

Intensive TA would be individualized and stakeholder specific. Using this framework, the modes of TA suggested by stakeholders include one-on-one training to help states get up and running with the system and immediate responses to queries.

Promote Peer Engagement

The needs assessment findings reiterate the importance of engaging stakeholders and promoting peer engagement. There was a particular focus on the FSS Data Center being “collaboration oriented.” There were a number of stakeholder groupings that could form “affinity groups,” such as state to other states, within and across agencies, between states and federal agencies, and researchers with practitioners, “programmers,” and administrators. Stakeholders noted that peer engagement is critical to sharing best practices, transferring knowledge, and helping systems to “talk to one another.” To encourage peer engagement, the FSS Data Center can:

- Facilitate relationships and partnerships between data providers and users
- Facilitate regional partnerships across states and localities for data sharing and analysis for common programs (e.g., TANF)
- Facilitate data sharing to enable states to share best practices with each other
- Provide a forum where data providers and users can share FSS research questions

Next Steps and Timeline

With the needs assessment complete, the project team, in consultation with OPRE, will continue to develop the framework for the Data Center while broadening discussions with states around their involvement in the Data Center. The Data Center will undergo a pilot year, which will be crucial to

testing the center's capabilities and troubleshooting glitches. It will also serve to make the Data Center more efficient for data users, as it will allow identification of any missing elements deemed critical by the pilot states. The pilot year may also identify any superfluous and redundant features and allow for those to be eliminated if necessary. During the pilot year, the Data Center will be refined to better address their needs as they arise throughout the year.