Expanding TANF Program Insights

A Toolkit for State and Local Agencies on How to Access, Link, and Analyze Unemployment Insurance Wage Data

TANF DATA COLLABORATIVE

OPRE Report 2022-226

OCTOBER 2022

Funders

The TANF Data Innovation (TDI) project is funded by the Office of Family Assistance (OFA) and the Office of Planning, Research, and Evaluation (OPRE) within the Administration for Children and Families (ACF). MDRC and subcontractors Actionable Intelligence for Social Policy at University of Pennsylvania, Chapin Hall at University of Chicago, and Coleridge Initiative are conducting the Promoting and Supporting Innovation in TANF Data project for the Department of Health and Human Services (HHS) under a contract with the Administration for Children and Families, funded by HHS under a competitive award, Contract No. HHSP233201500059I. The project officers are Nicole Deterding and Girley Wright.

Dissemination of MDRC publications is supported by the following organizations and individuals that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Annie E. Casey Foundation, Arnold Ventures, Charles and Lynn Schusterman Family Foundation, The Edna McConnell Clark Foundation, Ford Foundation, The George Gund Foundation, Daniel and Corinne Goldman, The Harry and Jeanette Weinberg Foundation, Inc., The JPB Foundation, The Joyce Foundation, The Kresge Foundation, and Sandler Foundation.

In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

FOR INFORMATION ABOUT MDRC AND COPIES OF OUR PUBLICATIONS, SEE OUR WEBSITE: www.mdrc.org.

Expanding TANF Program Insights

A Toolkit for State and Local Agencies on How to Access, Link, and Analyze Unemployment Insurance Wage Data

OPRE Report 2022-226
OCTOBER 2022

Authors:
Edith Yang, MDRC; Sharon Zanti, AISP; T.C. Burnett, AISP; Richard Hendra, MDRC; Dennis Culhane, AISP; Zarni Htet, MDRC; Della Jenkins, AISP; Camille Préel-Dumas, MDRC; Electra Small, MDRC

Submitted to:
Nicole Deterding and Girley Wright, Project Officers, Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Project Director: Richard Hendra, MDRC, 200 Vesey Street, New York, NY 10281.

Contract Number: HHSP2332015000591

Suggested citation: Yang, Edith, Sharon Zanti, T.C. Burnett, Richard Hendra, Dennis Culhane, Zarni Htet, Della Jenkins, Camille Preel-Dumas, and Electra Small. 2022. *Expanding TANF Program Insights: A Toolkit for State and Local Agencies on How to Access, Link, and Analyze Unemployment Insurance Wage Data*. OPRE Report 2022-226. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Overview

State and local leaders at Temporary Assistance for Needy Families (TANF) agencies have been increasingly focused on using administrative data from TANF and other state agencies to better assess how well programs are working, inform policies and practices, and, ultimately, improve the lives of families with low incomes. Economic mobility through employment retention and advancement is of particular interest to TANF leaders, but administrative data on TANF recipients' earnings are often difficult to access except for the purpose of investigating noncompliance.

Since 2017, the TANF Data Innovation (TDI) project has been helping state agencies harness their administrative data to improve family outcomes. Sponsored by the Administration for Children and Families within the U.S. Department of Health and Human Services, TDI is being led by MDRC in partnership with Chapin Hall at the University of Chicago, Actionable Intelligence for Social Policy at the University of Pennsylvania, and the Coleridge Initiative. This toolkit is part of that effort.

The aim of the toolkit is to offer practical guidance to state and local TANF agencies on how to access, link to, and analyze Unemployment Insurance (UI) wage data from state Departments of Labor for program monitoring, reporting, and evaluation. As such, the toolkit consists of a guidance brief and a companion GitHub repository:

The **guidance brief** is organized into four main sections: (1) a short **introduction** that lays out the purpose of the toolkit as well as background information on UI wage data and the kinds of research questions that data can be used to answer, (2) a description of common **challenges** to accessing state UI wage data and strategies to address those challenges, (3) methods for **linking UI wage data** to other data sources, including emerging advanced methods that are more secure, and (4) instruction for **preparing UI wage data for analysis**, including how to create common employment-related outcomes that the field has used for decades to measure employment trends, stability, and mobility.

The GitHub repository provides open source and accessible code for use with the fourth section of the guidance brief, described above. It includes code to use to look for common UI wage data issues and guidance on how to resolve those issues. In addition, documents in the repository walk users through a strategy for processing UI wage data to create an analysis file and employment-related outcomes of interest. Finally, the repository has a resources folder with related supplemental materials that have emerged from the larger TDI project as well as from the research team's meetings with members of an expert working group made up of researchers, policy professionals, and state and local TANF agency staff members that toolkit users may find helpful.

This toolkit is meant to be a starting point for TANF leaders who want to access and analyze UI wage data. It offers the essential building blocks you will need to get started on your data analysis journey. Supplemental materials in the toolkit's appendix, annotated bibliography, and GitHub resources folder can help you further. Let's get started!

Contents

List of Exhibits

TABLE

FIGURE

BOX

Acknowledgments

This toolkit was developed through an iterative, collaborative process between partners in the TANF Data Collaborative and members of an expert working group (EWG) made up of state and local TANF agency staff, researchers, and policy experts. The EWG met in November 2020 and April 2021 to provide expert guidance and real-life examples from their experiences in the field. The toolkit builds on decades of evidence on accessing, linking, processing, and analyzing UI wage data and represents a culmination of the joint energy and effort of these many contributors.

The authors are indebted to the EWG and the rich insights they generated to shape and improve the toolkit. Members include Michael Billingsley, California Department of Social Services; Yvette Chocolaad, National Association of State Workforce Agencies; Nicole Deterding, US Department of Health & Human Services Administration for Children and Families, Office of Planning, Research, and Evaluation; Dave Dlugolecki, New York Office of Temporary and Disability Assistance; Bob Goerge, Chapin Hall; Kathleen Moore, US Department of Health & Human Services Administration for Children and Families, Office of Planning, Research, and Evaluation; Demetra Nightingale, Urban Institute; Britany Orlebeke, New York Office of Temporary and Disability Assistance; John Padilla, New Paradigms Consulting, New Haven, CT; Kate Parr, University of Connecticut School of Social Work; Jeff Price, Virginia Office of Research and Planning; Daniel Schroeder, Ray Marshall Center for the Study of Human Resources, University of Texas at Austin; and Tayyab Walker, NYC Opportunity.

Additionally, the authors would like to thank Amy Hawn-Nelson and Deja Kemp at AISP; Barbara Goldman, Erika Lundquist, Caroline Morris, Stephanie Rubino, Johanna Walter, and Melissa Wavelet at MDRC; Bob Goerge, at Chapin Hall; and Julia Lane, at New York University for supporting and contributing to the EWG meetings. Girley Wright, Nicole Deterding, and Kathleen Moore, our sponsors and partners at the Administration for Children and Family's Office of Family Assistance (OFA) and Office of Planning, Research, and Evaluation (OPRE), as well as numerous staff members from the Division of Data and Improvement (DDI), also provided valuable guidance throughout the development of the toolkit. Part IV of the toolkit and the companion Github repository are drawn from decades of workforce data expertise and programming concepts developed by current and former MDRC staff, including Stephen Freedman, Gilda Azurdia, Mark van Dok, Victoria Deitch, Aaron Hill, Natasha Piatniskaia, and many others.

At MDRC, Johanna Walter provided a detailed external review of the toolkit and companion GitHub repository. Dannia Guzman and Anumita Jain provided crucial programming and technical assistance insights that helped to shape the toolkit's GitHub repository, and Rose Kob and Brence Pernell provided legal guidance on sharing open-source code. Jill Kirschenbaum edited the report and Carolyn Thomas designed and prepared it for publication.

The Authors

PART I
INTRODUCTION

The Purpose of This Toolkit

The goal of the Temporary Assistance for Needy Families (TANF) program, and social services more broadly, is to support the basic needs of families so that they can work, live, and thrive. In recent years, state and local TANF leaders, policy makers, and researchers have focused on how agencies can increase the use of TANF, employment, and other administrative data to better assess how well programs are working, inform policies and practices, and, ultimately, improve the lives of families with low incomes.

This toolkit, part of the TANF Data Innovation (TDI) project led by MDRC, was created to help TANF professionals develop more robust, data-driven practices using administrative data on earnings.[1] Whether you are a frontline case worker, a data analyst, or an administrator, the toolkit is designed to help you explore strategies to access and use earnings data for program improvement purposes.

Linking and analyzing TANF administrative records and state Unemployment Insurance (UI) wage data (the earnings data that are the focus of this toolkit) can help you answer key questions about client outcomes, program access and equity, and efficiency. For example, what percentage of your clients are employed five years after leaving the program and how much did they earn in the last year? How do employment outcomes of TANF leavers vary by race and ethnicity? How can your agency provide the most cost-effective supportive services to current recipients?

State UI offices frequently share wage data securely with third party research organizations to evaluate such metrics, demonstrating that linkage with TANF records is feasible if all requirements are met.[2] It is often difficult, however, for already-stretched TANF organizations to prioritize data work and analytics. A recent needs assessment by the TDI team showed that using employment information for data analysis related to programmatic improvement is not typical across TANF agencies.[3] Instead, those with access to UI wage data more commonly use them for investigating noncompliance within current caseloads.

1 Read more about the TDI project at https://www.mdrc.org/project/tanf-data-innovation#overview.

2 Recently, MDRC led the collection of UI wage data from 10 states for the SNAP E&T demonstration.

3 Goerge, Wiegand, and Gjertson (2021).

The primary purpose of this toolkit is to offer practical guidance to state and local TANF agencies on how to access, link, and analyze employment data from UI systems for program monitoring, reporting, and evaluation. The toolkit may also be useful to other state human services agencies (for example, SNAP and Child Support) that want to expand their data use, as well as policymakers interested in supporting improved workforce outcomes. State Department of Labor agencies may also gain useful insights from the data preparation section in Part 4, as well as from the broader discussion of ways to use employment data to improve human services programs.

Further guidance on TANF data use topics can be found in the GitHub repository, a companion to this toolkit that offers open source code and documentation for program administrators and researchers who are preparing employment data for analysis. Box 1 describes some of the terms that are used throughout the toolkit, to provide clarity for the different audiences that might use it. Relevant research examples are also linked to throughout the toolkit.

Box 1. Toolkit Terms

Linking vs. integrating vs. matching data: For the purposes of this toolkit, the word ***linking*** is used to refer to the technical process of merging two or more datasets. ***Integrating*** refers to a routine system of linking datasets that encompasses the technical aspects of data linkage as well as the ongoing governance structures. ***Matching*** is often done in conjunction with ***linking*** data but specifically refers to the process of a data provider or analyst using identifiers to find, or "match to," specific individuals in their databases.

Administrative data: Administrative data are data collected by government or other organizations for administrative purposes to document and track information such as monetary transactions, services provided, or participation in programs. The data are generally collected for the full population of relevant individuals. This toolkit focuses mainly on state Unemployment Insurance (UI) wage data, a source of administrative data on earnings. Within the toolkit, the term ***administrative data*** may also be used to refer to state TANF agency data on agency caseloads.

UI wage data: This toolkit focuses on accessing and using data from state UI agencies. Employers are required to report quarterly earnings of all their workers who are covered by their state's UI laws to the state Department of Labor, which manages and administers UI claims and benefits. These data generally cover more than 90 percent of employment in each state but do not include data for federal jobs, out-of-state jobs, self-employment, certain types of railway or agricultural jobs, or informal jobs.

(continued)

Box 1 (continued)

Employment: Throughout this toolkit, ***employment*** refers to an individual who has any earned income. Still, it must be recognized that using UI wage data to create employment measures may result in undercounting important sources of employment that do not show up as earnings in the state UI wage data files, and that self-reported information on the employment status of TANF recipients can be helpful in validating the accuracy of employment captured using administrative data.

What Are the Advantages of Using UI Wage Data?

Accessing UI wage data and linking that to TANF data has several advantages over using self-reported information on employment (which can include caseworker-collected information, survey responses, or pay stub verifications).[4] **First, historical UI wage data are collected for all individuals who earned income from employers subject to state UI laws,** while requests for self-reports (such as surveys) may not yield high response rates and that information is available only for specific points in time. UI wage data can be accessed for both past and current TANF recipients, so it is possible to perform data analyses on employment outcomes over time for both cohorts. **Second, UI wage data are more accurate and do not suffer from recall errors by respondents** because they are employer-reported and subject to audit and appeals from both employers and employees.[5] **Finally, self-reported data are susceptible to non-response bias**—that is, individuals with certain characteristics and life circumstances may be more likely to provide employment information than other individuals. For instance, people who experience employment instability may not want to disclose that information or may be difficult to reach for initial or follow-up survey completion. Gathering employment information by collecting UI wage data may help avoid such information gaps for hard-to-reach subpopulations and does not require new outreach to the clients.

What Are the Limitations of Using UI Wage Data?

Although UI wage data cover most employment in the United States, there are some limitations. UI wage records are reported quarterly, and most states do not have detailed information about hours worked or weekly or monthly employment. State UI data systems do not capture federal jobs, self-employment, out-of-state employment, and some categories of farm and railway employment. Federal wage data, such as the National Directory of New Hires (NDNH), tax data from the Internal Revenue Service (IRS), and the Social Security Administration (SSA) can capture some of the data that are not covered by state UI wage data. However, the barriers to accessing federal data are higher than

4 Dorsett, Hendra, and Robins (2018).

5 Barnow and Greenberg (2015).

those for state-level UI wage data and often require more intense security clearance processes. None of these administrative data sources capture informal or “off the books” employment.[6] For a comprehensive overview of the pros and cons of various self-sufficiency data sources, including TANF and UI records, check out the [Compendium of Administrative Data Sources for Self-Sufficiency Research].

Evidence shows that UI wage data may also miss more employment in populations with low incomes than those in higher-income groups, based on differential employment rates calculated from survey (self-reported) and administrative (employer-reported) data sources.[7] This means that the employment experiences of communities disproportionately living in poverty—particularly Black and Brown communities—may not be accurately captured with UI wage data. Additionally, three to six months of lag time is typical for UI wage data to be available. If an agency is interested in complete caseload employment data for 2021, for example, the data request could be made no earlier than the second quarter of 2022.[8] Nevertheless, for TANF agencies looking to make data-driven improvements in their programs, accessing and using state UI wage data may be the most practical and efficient way to perform employment-related data analyses.

Linking TANF and UI wage data can give both TANF and employment agencies a more complete picture of the trajectories of the people who access TANF services and allow those agencies to explore solutions to help workers achieve employment stability and economic mobility. For example, state TANF agencies need to understand the employment outcomes of TANF recipients after they leave or “exit” TANF, but TANF exit data are an unreliable measure of success.[9] Collecting UI wage data can provide a more complete sense of household earnings.

Linking TANF data with UI wage data may allow users to answer many valuable questions and help build a broader evidence-based agenda. Some examples are listed below. Box 2 discusses a specific example of how UI wage data have been used in evaluation research.

Integrating TANF data with UI wage data can help you:

- understand short- and long-term employment placement, retention, and advancement outcomes of TANF applicants, recipients, and leavers
- analyze trajectories and outcomes of TANF recipients who receive other services, such as childcare subsidies
- identify and compare segments of the TANF caseload, based on more complete data on prior experiences, that are more or less likely to have difficulty finding stable employment

6 Yang and Hendra (2018).

7 Abraham, Haltiwanger, Sandusky, and Spletzer (2013).

8 Yang, De La Rosa Aceves, and Tomlinson (2019).

9 Safawi and Pavetti (2020).

Box 2. Making a Difference with UI Wage Data

Many program evaluations, including those conducted by states, have incorporated UI wage data to better understand earnings outcomes and guide policy based on insights gained. One such example is MDRC's evaluation of [WorkAdvance](), an evidence-based program that helps participants prepare for and enter higher-paying jobs with better benefits and working conditions in sectors with opportunities for career growth and advancement. MDRC collected and analyzed UI wage data to examine the impacts and implementation of the program in Tulsa, New York City, and northeast Ohio, and found that WorkAdvance increased earnings over a five-year follow-up period and had positive benefit-cost findings. One site, the Per Scholas Institute of Technology in New York City, had among the largest employment impacts observed in random assignment evaluations of workforce programs, and its sector-based training model is now being replicated across the nation. The positive long-term benefit-cost findings relied heavily on the ability to analyze longitudinal UI wage data on study participants.

The impact findings from WorkAdvance influenced important legislation such as the Workforce Innovation and Opportunity Act reauthorization and the SECTORS bill. WorkAdvance was also the only job training model named as a promising means of reducing child poverty by the National Academies of Sciences, Engineering, and Medicine. Thus, the evaluation had a far-reaching influence on federal, state, and local policymaking. The findings would not have been possible without UI wage data.

- describe labor market characteristics of past or current TANF recipients, such as prevalent industries and respective wages and retention outcomes
- understand how the policy and economic context—at the ZIP code or census block level—influence employment outcomes for TANF participants
- evaluate two-generation outcomes of children and their parents who received TANF[10]
- explore characteristics of employers who successfully hire and retain TANF recipients
- create new tools, like a dashboard that tracks self-sufficiency outcomes, to help better manage program performance[11]

---

10 Two-generation programs approach the well-being of the whole family and look at how interventions might improve circumstances of both parents and children. See Chase-Lansdale and Brooks-Gunn (2014) for more detail.

11 Additional analyses, such as mapping geographies of employment to look at accessibility for TANF recipients, may also be possible with non-state sources of administrative earnings data, such as On the Map.

Bringing a Diversity, Equity, and Inclusion Lens to Working with Data on the TANF Population

While there are challenges to accessing other administrative data sources and linking those data to TANF records, the insights gained can lead to more effective program and policy development. However, administrative data alone cannot provide a full picture of client experiences and should not be your only tool for better understanding outcomes and issues like program access and equity. It is important to bring in additional sources of information that can further contextualize the data, findings, and implications, such as analyses of root causes, descriptions of the employment and political climate, or your program's history.[12] Incorporating qualitative stories and the perspectives of former or current TANF recipients, although challenging to collect, is a valuable way to contextualize findings. Consider including people with lived experience with TANF or other relevant public systems and services on advisory councils or in other community engagement forums and be sure to compensate them for their time.[13]

Access and outcomes within TANF are disproportionately distributed by race: Black families are more likely to live in states with the lowest TANF benefits and to experience more employment instability after exiting TANF; Black and Hispanic participants are more likely to be sanctioned than White participants; and White families tend to get offered more supportive services during their TANF participation.[14] While there are numerous factors at play in these trends such as poverty, geography, and access to education, race is particularly significant. Black, Indigenous, and people of color (BIPOC) in the United States, regardless of socioeconomic status, tend to experience worse outcomes across numerous social service system measures.[15] Issues of racial bias may also appear in data from UI systems. For example, as previously mentioned, UI wage data may miss employment information for BIPOC families at higher rates than for White families. Another key source of UI wage data, the National Directory of New Hires (NDNH), was originally compiled to enforce child support court orders and may reflect the biases embedded in the enforcement of that program.

Bringing an equity lens to accessing, linking, and analyzing TANF and UI data requires doing more than reaffirming these inequities: It means taking steps to examine the structural racism and policies that caused these inequities, and to identify opportunities for change. Administrative data can be a valuable tool to support this process. AISP's Toolkit for

12 See Aid to Dependent Children: The Legal History for the history of TANF and its predecessor, Aid to Families with Dependent Children.

13 See AISP's working paper, Addressing Racial and Ethnic Inequities in Human Service Provision, for more guidance on how to assess program equity and create a plan of action to correct for inequity.

14 Center on Budget and Policy Priorities (2022); McDaniel, Marla, Tyler Woods, Eleanor Pratt, and Margaret C. Simms (2017).

15 McDaniel, Marla, Tyler Woods, Eleanor Pratt, and Margaret C. Simms (2017); Hayes-Greene, Deena, and Bayard P. Love (2018).

Centering Racial Equity Throughout Data Integration offers guidance on how to incorporate an equity lens at each stage of the administrative data life cycle. In addition, AISP has developed a list of concrete ways to bring a race equity lens to projects using TANF and UI wage data, available in the Github repository. More resources can be found in the Appendix.

PART II
DATA SHARING CHALLENGES AND STRATEGIES FOR SUCCESS

Barriers to Accessing and Linking UI Wage Data

Although every TANF agency is unique and will start its data sharing journey from a different place, there are several areas where challenges commonly arise when working with UI wage data.

Real or Perceived Legal Constraints

Any time agencies share and link data at the individual case or person level, comprehensive legal agreements must be put in place to protect client privacy and to ensure the data are used legally and ethically.[1] This inevitably requires the agencies to work together to determine how the data will be used and to meet all of the legal requirements associated with those uses. Such negotiations take time, particularly if data users are siloed within their agencies or do not have clear directives from leadership for how to engage in the collaboration. The real or perceived threat of legal risk or regulatory enforcement can halt or delay access, even when there is conceptual agreement among all of the partners on the benefits of data sharing and staff time has been budgeted to support such work. In TDI interviews, agency staff reported that this apprehension stems, at least in part, from uncertainties regarding the tangle of state and federal privacy laws that apply to the linkage and use of TANF and wage data as well as concerns on the legal side about the specifics of how data will be used. Less often, but perhaps more frustrating, legal questions may be raised as an automatic, risk-averse response to data sharing requests.

1 See OPRE's confidentiality toolkit for data interoperability for more on relevant privacy concepts and laws. Additional recommendations on responsibly sharing confidential data can be accessed at https://www.acf.hhs.gov/opre/project/responsibly-sharing-confidential-data-tools-and-recommendations.

Data Access

State and federal wage and income data are critical to tracking employment and earnings over time, yet they are among the more challenging datasets to access. State UI systems and the IRS are the two primary sources of U.S. earnings data. Each state has discretion over the operation of and rules for its own UI program and Department of Labor (DOL), so the parameters for data access vary accordingly. Some states have standard policies and procedures for linking UI wage data with education and training program records, for example, but these governance processes have not often been extended to linking UI and TANF data.[2] In many states where linking UI wage data with other sources is allowed, there may be restrictions on *when* it can occur. For example, administration or performance measurement may be allowed, but using the data for research and evaluation may be more restricted. Cost is also an important consideration: Because providing data to another agency is often not part of a given UI agency's budget, there may be administrative fees for accessing UI wage data files—for example, to account for programmer time or costs per record. Identifying funding to support TANF and UI wage data linkage can help overcome stalled negotiations.

State wage data access challenges are only heightened during economic downturns. That was the case early in the COVID-19 pandemic, when each state's DOL was tasked with responding to an increased number of UI claims. Accessing federal data, such as those compiled through the National Directory of New Hires (NDNH) and the Longitudinal Employer-Household Dynamics (LEHD) program, among others, can also present challenges. The approval processes necessary to access these sources and the data lag associated with some of them may limit their usability for time-sensitive analysis and decision-making. Additionally, the NDNH in particular has limited historical data and constraints on analysis procedures, such as restrictions on small sample sizes and pre-specification of variables to be included in the analysis.

Capacity and Resources

Staff time and analytic expertise are often limited in both TANF and state DOL agencies. The TDI needs assessment showed that busy and underfunded TANF agencies must continually prioritize their most important, urgent data requests. This often means that data analysis beyond what is required for standard federal reporting purposes or state management is out of reach, making it difficult to share data across agencies or with third parties such as universities and research organizations. Even when long-term efforts to share and integrate data are initiated, they often get put on the back burner in favor of more immediate tasks. Negotiating access to wage data, wrangling the data, and conducting sophisticated analyses requires substantial investment in staff time and staff training. Moreover, it takes time to build the cross-agency relationships necessary to facilitate ongoing data sharing and to cultivate those connections over time. Without these

2 See the U.S. Departments of Labor and Education's Joint Guidance On Data Matching To Facilitate WIOA Performance Reporting And Evaluation for an overview of the multiple ways in which education and wage data can be linked and the various legal auspices under which these data can be shared with analysts.

investments in relationships, training, and time, it can be difficult for agencies to retain historical knowledge about complex programs and data systems, plan for staff succession, and build a sustainable model for ongoing data access.

Technical and Analytic Limitations

In general, the reuse of administrative data for research and evaluation purposes comes with its own distinct challenges, given that the data were originally collected for service delivery, compliance, and other non-research-oriented needs. For example, TANF and UI wage data are primarily collected for determining eligibility for benefits. Missing data, limited access to shared identifiers (for example, having only Social Security numbers), and a lack of a historical record on variable changes over time can all hinder linkage and analysis. Legacy data systems may be antiquated and prevent data from being easily extracted, and the transition to new systems can change access and linkage procedures.[3] Wage records generally do not include details on job supports provided or occupations represented and may have no information on workers whose jobs are not covered by UI. Data storage may also be a concern among agencies that are amassing ever-larger administrative data sets. Finally, data security presents legal concerns as well as technical challenges in implementing robust security plans within budget constraints. These issues are compounded when analytic questions require data access and linking across states, which is often necessary in regions where workers frequently cross state lines for employment.

Cross-Agency Relationships and Politics

Even with the best technology platforms, analytic capacity, and legal processes in place, data sharing between TANF and DOL agencies is difficult to accomplish without strong cross-agency relationships and communication pathways. It can be particularly challenging to get key working partners on board if there are prior trust issues between agencies, heightened concerns around auditing and liability, or just a lack of familiarity with the mutual benefits such a partnership may bring. Changes to agency leadership and shifting political tides can also affect the ease of and support for data access. If support for data sharing partnerships fades with administration changes, it can be difficult to maintain momentum. Furthermore, insights from integrated data will be much more useful if the agency has built an organizational culture of using data for quality improvement rather than solely for compliance purposes.

Identifying which of these barriers are at play in your specific context can be a helpful step toward diagnosing what next steps are needed.

Building Blocks for Successful Data Sharing

The following strategies can help you address common challenges to data sharing. Again, every agency is unique, so approaches should be tailored to your local context.

3 Evermore (2020).

Cultivate Cross-Agency Partnership and Trust

Data sharing is as relational as it is technical and moves "at the speed of trust."[4] If your agency is interested in accessing wage records and linking with TANF or other administrative data, engage with potential partner agency staff members early and often. Get to know the names, titles, and contact information of the key data gatekeepers at the state DOL agencies. Learn about what's important to them, what *their* challenges are, and what factors could motivate them to share data. Try to understand the reasoning behind any hesitancy or objections to sharing data: Earnings data are highly sensitive, for example, and UI agency staff members may be especially cautious about sharing that information. Forming a coalition of agencies invested in data sharing and identifying champions at each agency can help build stronger partnerships. Emphasize shared learning opportunities that aim to strengthen and improve programs rather than expose deficits. Relationship and trust building rooted in a shared understanding of opportunities and risks can help overcome the common fear that negative results will be presented out of context and lead to negative publicity for agencies.

Align Priorities and Goals

Identifying issues in which program administrators, elected officials, and working partners have a shared interest can be a way to gain support for both improving data infrastructure and prioritizing data access. Consider mapping exercises in which all entities involved in the data sharing enterprise highlight what they stand to contribute and gain from the arrangement. High-level priorities shared by TANF and DOL agencies may include improving economic security, economic mobility, and racial equity, as well as reducing disparities in access to services and outcomes and identifying earnings trajectories of TANF recipients. Labor and workforce development agencies specifically may support data sharing if it improves their understanding of the TANF recipients they serve and the employers who hire them. Some jurisdictions may find it helpful to form a council of agency leaders around a shared policy or program agenda. This creates a regular opportunity for exchange and collaboration across agencies that need to share data. In addition, getting executive leadership involved in such efforts may speed cooperation. Particularly during times of crisis, it may also be helpful to prioritize linking TANF and wage data to facilitate a rapid response to basic needs as well as the longer-term economic recovery.

Develop Cross-Agency Data Governance

Policies and procedures that specify the data sharing "rules of the road" help build institutional trust and ensure that necessary safeguards and protections are in place so all partner agencies contributing and receiving data can fulfill their roles as data stewards. Governance also helps ensure that data are used ethically by implementing collaborative and transparent processes to resolve questions about appropriate analytic techniques, data quality, bias, and other issues. Finally, data governance supports data sharing efforts during political shifts and helps build long-term sustainability. In some states, TANF and DOL

---

4 Hawn Nelson et al. (2020).

agencies may benefit from using existing cross-agency data governance capacity such as those provided by Integrated Data Systems, statewide Longitudinal Data Systems, or P-20 Data Systems. These data sharing infrastructures may already have strong processes in place and can serve as neutral third parties to facilitate linkage. Such models are discussed in more detail in Part III.

Table 1 outlines common data governance activities and the various working partners you will want to engage. You can also find more resources for building a governance framework in the Appendix.

Table 1. Data Governance Activities Throughout the Project Lifecycle

GOVERNANCE ACTIVITIES	WHO SHOULD BE AT THE TABLE?
Prioritizing shared inquiry questions	Executive leadership, all data partners
Agreeing on a data model, assessing data availability and quality (see Part IV for more on data models)	Agency staff from all data partners, including analysts and those closer to data collection and data management
Ensuring ethical data use	Executive leadership, all data partners. May also involve external ethics review and/or IRB
Developing and executing legal agreements	Legal counsel from each data partner agency, privacy officers, CDO/CIO
Monitoring data security	IT support, data managers, privacy officers
Analysis and interpretation	Agency staff conducting analysis along with research partners from partner agencies and institutions as needed
Translation and use	Executive leadership, data partners, and data consumers

Develop a Legal Framework

A key step toward “getting to yes” is to provide the legal departments of all partnering agencies with granular detail about the analytic project, including its purpose; how access, linkage, and analyses will be performed; and how data will be released. Providing this information upfront can help teams more quickly and easily determine whether or not the use of data is legal. As partner agencies develop a shared understanding of the structures and processes that will govern and enable data integration, their legal counsels should document these processes in legal agreements along with rigorous data security requirements. It may also accelerate future negotiations to develop a standard process and templates for legal agreements. (See Box 3 for a description of the types of agreements you

Box 3. Understanding Legal Agreements

The names of the legal agreements that facilitate the sharing of data vary across contexts but generally speaking, there are three agreements that are commonly needed for any data-sharing effort:

1. A memorandum of understanding (MOU) outlines the broad purpose, parties, terms, data request process, and authority of the signatory.
2. A data sharing agreement (DSA) works in tandem with an MOU. It outlines the respective legal rights and responsibilities of each party using identifiable data for linkage for approved uses.
3. A data use license (DUL), also referred to as a data use agreement (DUA), governs the release of data that have been de-identified for analysis and is signed by the data user for each approved project.

In summary, the MOU is often a broad document, signed by all data partners. The DSA is a technical agreement, specific to identifiable data being used for linkage—the agreement for data used for integration. The DUL or DUA is specific to an approved use of data for a specific purpose—the agreement for data used for analysis.

will likely need.) Your partner agencies may use different terms to refer to these documents. Regardless of what they are called, it is important to be clear about the purpose of each agreement within the larger framework as well as to learn your partner's preferred terminology and to use these terms consistently. Also keep in mind that how you develop legal agreements (including the types of agreements you use, who should sign them, what's included, and so on) depends on your method for linking TANF and UI data. See Part II for a high-level overview of legal considerations for each data linkage method.

Understanding the legal limitations and requirements to which partner agencies must adhere may be helpful in navigating legal barriers. For example, employment agencies may have specific language to describe the various purposes for which data sharing is or is not allowed; learning this language and incorporating it into legal documents can speed progress. Similarly, legal difficulties may be eased by identifying the "what's in it for them" factor and ensuring all data sharing entities are deriving value from the exchange. Negotiating legal agreements takes persistence. You have to establish trust among all parties and collaboratively weigh the risks and benefits of data sharing.

After agreements have been executed, it is of the utmost importance that agencies adhere to all legal obligations, training requirements, data retention and destruction protocols, and other contractual requirements to protect privacy, maintain agency ethics, and solidify trust in data sharing going forward. The Appendix offers additional resources to help you navigate legal barriers and execute data sharing agreements.

Develop a Data Security Plan

Data security is often categorized as a technical consideration, even though it is a multidimensional process that includes legal, technical, procedural, and physical components. The goal of these multiple layers of security is the same: to prevent a data breach or security incident. While catastrophic events are uncommon, one-off errors involving an individual client's information are more common. So it is important to consider whether your project requires sharing personal identifiers or if other options are available to avoid exposing these sensitive data. For example, that could mean using hashed IDs, which are scrambled ID numbers to protect identifiers, or sharing data through a separate secure process such as an identity resolution center, which uses anonymized IDs and other systems data to link separate data sources. Sharing how an agency prepares for and responds to data security threats is essential for building and maintaining trust with data partners and the broader community. As such, it is imperative that you develop a clear incident protocol to prepare for potential data breach and security incidents. See the GitHub repository's Resources page for a self-assessment you can use as a starting point to gauge your organization's data security posture and identify potential next steps for enhancing capacity in this area. Note that this self-assessment is based on the National Institute of Standards and Technology 800-53 standards and may not cover every requirement in a state or locality.

Table 2 offers a menu of legal, technical, procedural, and physical conditions to consider when addressing data security.[5] It is important to include elements from each dimension in your data security plan, which may be articulated in legal agreements or policy documents. Every state or local agency will have its own configuration of interlocking data security components, in addition to any federal requirements that apply. Data security plans or self-assessments should take those requirements into account.

Document the Historical, Legal, and Technical Milestones

While documenting any changes to data or security practices may be required by legal agreements, it is also critical to document changes to the data sharing context and governance process over time. Record the history behind the data sharing effort, how it started, how it has developed, and key milestones (such as signing MOUs or other legal agreements or passing legislation or executive orders that enable data sharing). This will make leadership and staff transitions smoother, build trust with partners, and increase the continuity and sustainability of data sharing. Similarly, on the technical side, providing detailed documentation within any programming code on the reasons for making data processing or analysis decisions is highly recommended to help current and future analysts track changes to data and methodologies over time. This also helps to ensure that data security safeguards that were required in any executed agreement are maintained by staff members working directly with the data.

5 Hawn Nelson et al. (2020).

Table 2. Legal, Technical, Procedural, and Physical Conditions to Consider When Addressing Data Security

LEGAL	TECHNICAL	PROCEDURAL	PHYSICAL
• Documentation of signatory authority • Memoranda of understanding • Data sharing agreements with data partners • Data use agreements or licenses with data users • Confidentiality agreements for individuals who have access to identifiable data	• Regular security audits • Digital access controls (multi-factor authentication preferred) • Encryption – data at rest, data in transfer • Secure servers • Data integrity measures (such as backups) • Controlled, limited access • Private network • De-identification guidelines	• Data governance board or agency oversight • Logs/audit trail • Collaborative checklist for data requests • Regular communication between staff • Documentation of business process (to explain "how we work") • Staff training (including annual review of confidentiality agreements) • Incident response protocols (such as what happens in the event of an incident or breach) • Clearly documented separation of staff duties	• Safe (for storage of physical data, if applicable) • Locked offices • Hardened work stations (devices secured to mitigate unauthorized access and use)

SOURCE: Amy Hawn Nelson, Della Jenkins, Sharon Zanti, Matthew Katz, T.C. Burnett, Dennis Culhane, and Katie Barghaus, *Introduction to Data Sharing and Integration* (Philadelphia: Actionable Intelligence for Social Policy at the University of Pennsylvania, 2020). Website: https://aisp.upenn.edu/resource-article/introduction-to-data-sharing-and-integration/.

Approach Data Sharing in Phases

Thinking about a data sharing partnership in phases can be a strategic way to generate small wins and maintain momentum for the larger data sharing effort. Such an approach might start with building shared agreement on priorities, developing governance processes and an overarching legal agreement, and then executing data use agreements as needed for specific projects. Another common step for advancing data sharing efforts is to get key leaders onboard. Once small successes are demonstrated, these leaders may even become champions of the effort. A phased-in approach can help data partners see how priorities, research questions, technology, access procedures, and other components will shift over time and allow for course correction at each step of the process.

Design and Document a Thoughtful Data Processing Approach

Once data access is established, agencies should only seek to link data elements that are both relevant to the analytic question at hand and of sufficient quality to provide insight; no one benefits from creating an elaborate dataset that is not used. Agency partners should agree on standards for data quality and then establish strategies by which quality may be built into the processes of measurement, collection, record transfer, and analysis. This will require continued exchanges between those who are building the data sets (generally, TANF and DOL analysts and technologists) and those who work most closely with the data day-to-day (generally practitioners and program staff).[6] Section IV of this toolkit provides technical details on preparing UI wage data for data integration efforts.

Build Staff Capacity and Expertise

Investing in your staff's capacity through training, resources, and team development will not only benefit short- and mid-term data projects but also allow you to better strategize and plan for succession over the long-term. Prioritizing staff training up front may also help avoid future confusion over roles and responsibilities and mitigate any loss of institutional knowledge resulting from staff turnover. See the Appendix for training and capacity building resources such as the Applied Data Analytics program available through the Coleridge Initiative.

6 Wiegand et al. (2017).

PART III
KEY METHODS FOR LINKING TANF AND UI WAGE DATA

As described in Part II, TANF and UI wage data can be linked legally and securely at the individual case or person level, though this process is not without its own set of challenges. This section lays out six key methods for linking these data. Deciding which one is best for you depends on your agency's goals, partners, and local context. For example, continuous data matching through an external, centralized data linkage center may be better suited to ongoing program measurement, whereas working on an ad-hoc basis with a third party such as a research organization or a university may be the right course for a particular program evaluation. There are pros and cons to each method, as described below, but all can be used to securely link and analyze TANF and wage data. The method you choose should take into consideration any resource constraints or existing relationships with partner organizations. And remember: You don't have to start with something complex to be successful. As discussed earlier, approaching data sharing in phases (perhaps starting with one method until another becomes feasible) is recommended.

The key linkage methods are defined as follows:

- **One-way data sharing.** A state or local TANF Agency shares TANF data with a state Department of Labor (DOL) office for linkage and analysis. Analyses may be completed by DOL staff or the DOL may invite TANF agency staff to complete analyses on-site.

- **Two-way data sharing.** A state or local TANF agency shares selected person-level identifiers such as Social Security numbers to request UI records for specific individuals, and the state DOL office in turn shares wage, UI, or similar records with the TANF agency for analysis. Linked records may or may not include the identifiers, depending on the agreement and project needs.

- **Sharing with a third party for analysis.** A state or local TANF agency and a state DOL office both share their data with an outside research organization or other third party group that conducts the linkage and analysis. This is particularly beneficial when the third party has both the analysis expertise and the security infrastructure needed to handle the data. For examples, see MDRC's Wage Data Study with Change Capital Foundation grantees and the work of the Administrative Data Research and Evaluation (ADARE) alliance.

- **Sharing in a data linkage center.** A state or local TANF agency and a state DOL office both share their data in a centralized system—commonly referred to as an *integrated data system* (IDS)—that has a governance, legal, and technical framework for securely and ethically linking and analyzing the data. Data are often linked based on common identifiers or hashed (anonymized) IDs. For some state agencies, this arrangement may provide a more neutral and trusted platform for routine data sharing. Data linkage centers come in many shapes and sizes. Some have been built by large public agencies or executive offices in government (such as Washington State's Integrated Client Database and Iowa's Integrated Data System for Decision-Making) while others are operated by university partners (such as the Administrative Data Research Facility at the Coleridge Initiative and the California Children's Data Network). Many data linkage centers offer remote access to linked data for approved users (via authenticated VPN), so TANF agencies may have the option to conduct their own analyses of the data or lean on the analytic capacity of the center staff.

- **Group linkages.** A state DOL office shares aggregate wage records and other summary statistics for pre-specified groups of individuals with a state or local TANF agency for analysis. This approach does not require person-level record linkage, which can be helpful for overcoming concerns about using individual-level data, as demonstrated in several evaluations, such as MDRC's Opening Doors Project, the Santa Clara Moving to Work evaluation, and the Work Advancement and Support Center Demonstration (Bridgeport).

- **Privacy-Preserving Record Linkage (PPRL).** PPRL is an umbrella term for record linkage approaches that involve linking case- or person-level records across secure databases maintained by different organizations without using sensitive identifiers such as SSNs or names.[1] Leading edge PPRL approaches, which go by different terms such as "secure multiparty computing" (SMPC) and "secure hash encoding" (SHE), are not yet common practice for data linkage, though some agencies have started implementing them. For example, the Department of Human Services in Allegheny County, Pennsylvania, conducted a demonstration project using PPRL to analyze human services data. Similar efforts have been undertaken in Chicago and Tulsa. These approaches excel at protecting sensitive information that is stored in several different systems by keeping it encrypted and in place in its original system while facilitating analysis.

Table 3 presents advantages, limitations, and legal considerations for each of these linkage methods. While reading this table, keep in mind that the information is framed from the perspective of TANF agencies. Other agencies and organizations may have somewhat different perspectives regarding the pros and cons of each method.

1 Randall, Brown, Ferrante, and Boyd (2019).

Table 3. Matrix of Key Methods for TANF Agencies to Link Data with UI Wage Records

	ONE-WAY DATA SHARING	TWO-WAY DATA SHARING	SHARING WITH A THIRD PARTY FOR ANALYSIS	SHARING IN A DATA LINKAGE CENTER OR INTEGRATED DATA SYSTEM	GROUP LINKAGES	PRIVACY-PRESERVING RECORD LINKAGE (PPRL)
Pros	Often the path of least resistance	Allows analytic flexibility	Third party such as a university or a research organization may lend additional capacity and analytic expertise Neutral ground for analysis	Greatly eases access and reduces time associated with approving and executing projects because data governance and access procedures have already been standardized Generally higher data quality and documentation Neutral ground for analysis	Removes individual identifiers, for agencies concerned about revealing person-level earnings Provides mechanism for matching when person-level matching arrangements cannot be negotiated	Security, governance, compliance requirements may be easier to meet because limited access to individual-level identifiers greatly reduces risk Utility of data remains high because linking and analytics can still be performed at the individual level

(continued)

Table 3 (continued)

	ONE-WAY DATA SHARING	TWO-WAY DATA SHARING	SHARING WITH A THIRD PARTY FOR ANALYSIS	SHARING IN A DATA LINKAGE CENTER OR INTEGRATED DATA SYSTEM	GROUP LINKAGES	PRIVACY-PRESERVING RECORD LINKAGE (PPRL)
Cons	Often only gets aggregated results, limiting further analysis Typically a one-time effort, not ongoing Less control over producing accurate results Typically requires new agreement for each project unless an SOP is put in place	Substantial time associated with getting DSA Typically requires a new agreement for each project unless a SOP is put in place	Requires DUA/DUL which could be time consuming Typically a one-time effort, not ongoing	Must address concerns about data being outside direct physical control of the DOL office Must identify existing linkage center that meets the needs of partners and is trusted as an intermediary Establishing a new data linkage center requires large investment of time and agency commitment	Requires significant advance thinking about statistical analysis and how group cells are created Lack of analytic flexibility, such as the limited ability to do analysis in cohort variation Complexity of methods means higher risk of error	Privacy-preserving technologies still evolving and require a specialty technology provider

(continued)

Table 3 (continued)

	ONE-WAY DATA SHARING	TWO-WAY DATA SHARING	SHARING WITH A THIRD PARTY FOR ANALYSIS	SHARING IN A DATA LINKAGE CENTER OR INTEGRATED DATA SYSTEM	GROUP LINKAGES	PRIVACY-PRESERVING RECORD LINKAGE (PPRL)
Legal considerations	Requires a MOU between two agencies that meets security, access, and analytics standards	Requires a MOU between two agencies that meets security, access, and analytics standards	Requires a DUL or DUA with the third party, possibly with both agencies (could be multiple agreements)	Requires multiple agreements: an MOU to set policies and procedures for data access and use and separate DULs or DUAs to authorize individual projects, but these agreements may already be templated for ease of negotiation.	Requires a DUA, but may have fewer data security requirements	Generally decreases the risk of data sharing with a third party in a manner that speeds up compliance and governance processes and reduces the need for some types of legal agreements (because of limited access to individual-level identifiers)

NOTES: DSA=data sharing agreement, DUA=data use agreement, DUL=data use license, SOP=standard operating procedure, MOU=memorandum of understanding.

Many states have found two-way data sharing to be a straightforward method for linking and analyzing TANF and UI wage data that allows for analytic flexibility. If partners are not yet willing to share individual-level data, some jurisdictions have found it helpful to request aggregate data through a one-way data sharing agreement. While this approach is not ideal, it is sometimes a helpful starting place and can be used to highlight the reasons for pursuing individual-level data linkage. Additionally, utilizing an IDS or other data linkage center as a trusted intermediary can be helpful due to built-in governance and legal frameworks. Other potential trusted third parties include university and research firms with capacity to support individual linkage or group match techniques. When there are significant concerns about data security that cannot be addressed through other methods, it may be worth exploring more technologically advanced approaches like PPRL. Regardless of the linkage method you choose, it is important to uphold all data security requirements and make sure your partners are comfortable with the various data linkage procedures. This will promote trust and support ongoing partnerships.

PART IV
PREPARING UI WAGE DATA FOR ANALYSIS

The previous sections of this toolkit examined the advantages of using UI wage data and its limitations, described data sharing obstacles and how to overcome them, and outlined the key methods for linking TANF and UI wage data sets. This section moves you on to the final step: preparing the data for analysis. It includes how to create a UI wage data request file, what to do once you have secured an agreement to access the data, how to conduct data quality checks, how to process the data and create employment-related outcomes, and how to link TANF and UI wage data for analysis.[1]

The technical guidance in this section is aimed at TANF agency staff members who are comfortable with coding and learning new programming languages but who may be new to working with UI wage data. Supplementing this section is the GitHub repository of hands-on computer code and documentation to help you check and process the data, including: (1) R markdown files that can be adapted for common UI wage data quality checks, (2) synthetic UI wage data files already programmed with selected data quality issues so you can write your own quality checks and test your code, (3) SQL code that can be used to process data and create outcomes, and (4) scripts to produce synthetic TANF cross-reference and UI wage data files you can use to test out your own code.

The process described in this section assumes the following scenario: A state TANF agency wants to explore employment outcomes of adult TANF recipients in the three years after their participation in a particular program. The agency has a two-way data sharing agreement with the state UI agency that allows for matching Social Security numbers (SSNs) for TANF families and for the UI agency to share back individual-level quarterly wage records to cover a specified period of time. In this scenario, the TANF agency sends a sample definition file (called a "request" file) with a list of SSNs. The UI agency matches those SSNs to the wage data in its system, and then transfers the UI wage data with SSNs back to the TANF agency. Of course, depending on your organization's specific goals and context, your scenario may not be identical to the one laid out here. But the tools provided in this section

1 This toolkit does not address how to prepare TANF data for analysis. For more information on this topic, see Wiegand et al. (2017).

(and in the GitHub repository) can still help you understand how to clean and restructure UI wage data and create commonly used employment-related measures for analysis.

Requesting Data from the State UI Agency

Depending on your linkage strategy, the usual first step in requesting wage data from a state UI agency is to create an **individual-level request file** with identifiers (SSNs) that the agency will use to extract a file with those individuals' wage records. Since TANF data are case-level data and UI wage data are individual-level data, it is important to consider which adults in each TANF case should be included in the request file. (That is, whose UI wage records do you want?) While most TANF cases are single-adult households, there are other possible case configurations. Sometimes there are multiple adults in one case, or child-only cases that have no adult recipients. Sometimes adults switch from one existing case to another, and sometimes the case ID itself changes over time. If you are interested in employment outcomes for household heads only, for example, you'll want to make sure to include the SSN for the correct adult in multiple adult cases. If you are interested in employment outcomes for all adults, you will want to include all SSNs and you will end up with more than one row on the UI wage data request file for multiple adult TANF cases.

It is also important to preserve a **TANF cross-reference file** with both case and individual identifiers, so that the returned data can be linked back to the correct individual in each case once the data are checked, processed, and transformed. Figure 1 shows some rows from a TANF cross-reference file with examples of different types of cases that would be converted to one list of SSNs for the UI agency. (Note that this process may be slightly different if you do not have a two-way data sharing agreement).

If you are tracking cases over time, it would be useful to include in your cross-reference file case IDs as snapshots over time, so you can accurately link earnings data back to the correct cases if these nuances are relevant for your analysis question. In Figure 1's example:

- There are six cases in 2018 and five in 2019, for six adults and one child.
- SSN 222222222 was a single-adult case in 2018 but joined case 00001 in 2019.
- SSNs 333333333 and 444444444 are two adults on the same case in both years.
- Case 00006 is a child-only case, and there is no SSN associated with it. Sometimes dummy SSNs are used in these cases (such as 999-99-9999). (Note that in an employment outcomes analysis, depending on your analysis question, it may be appropriate to drop child-only cases from the analysis sample since children are not of working age.)

Once the request file is created, it can be securely transferred to the UI agency using the agreed-upon method in your memo of understanding. The UI agency will match the identifiers to the specified time frame of your request and securely send a matched wage records data file back to you.

Figure 1. Creating a Request File from TANF Data

TANF CROSS-REFERENCE FILE

Case_ID_2018	Case_ID_2019	SSN	ProgStartDate	ProgEndDate
00001	00001	111111111	01/15/2019	01/15/2020
00002	00001	222222222	01/15/2019	01/15/2020
00003	00003	333333333	04/15/2019	01/15/2020
00003	00003	444444444	04/15/2019	01/15/2020
00004	00004	555555555	07/15/2019	
00005	00005	666666666	09/15/2019	03/15/2020
00006	00006		01/15/2019	

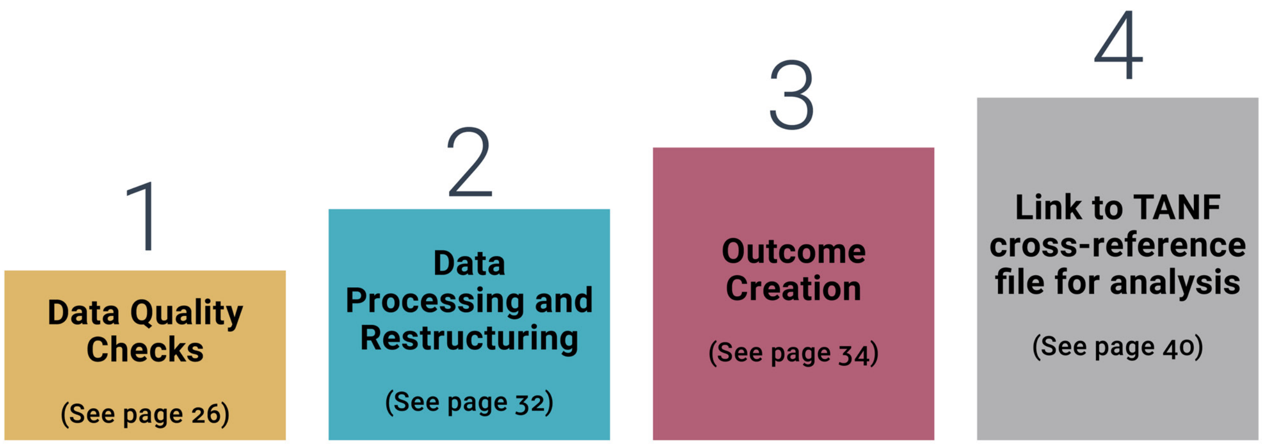
UI WAGE DATA REQUEST FILE

SSN
111111111
222222222
333333333
444444444
555555555
666666666

Getting Started on Preparing UI Wage Data for Analysis

Once you have received the requested UI wage data file from your state UI agency, there are a series of steps to complete before starting your analysis. Figure 2 illustrates this process, and also refers you to the pages in this part of the toolkit if you are looking for specific guidance on a particular step. The first step is to conduct data quality checks to make sure that the match to UI wage records was done correctly, and that the proportion of adults who show up in the file as having earnings, as well as the earnings amounts in the file, fit within your expectations. Once you are fairly confident that the data in your UI wage data file are accurate, you can process and restructure the data to create an analysis file. Next, you can create your outcomes of interest. The outcome creation section below walks through creating intermediate outcomes that make it easier to create field-tested employment-related outcomes that are used to measure any employment, employment retention and stability, and advancement. Finally, you can link this file back to your TANF cross-reference file to perform your analysis.

Figure 2. Sequential Steps for Preparing UI Wage Data for Analysis



Data Quality Checks

Performing quality checks on your data is a critical part of the analytical process. This step includes identifying and resolving common UI wage data problems that occur when receiving, restructuring, processing, or creating outcomes from your data. It is important to understand the context in which these data are collected to assess whether the data received from the UI agency are of good quality, meaning they are complete, consistent, understandable, and contain reasonable values. Like all data that are entered into a system, some entries will be incorrect due to human error when entering the information. Additionally, if employers report incorrect data and do not fix the errors within a certain time period, the erroneous records will remain in the state database. Considerations include:

- Do the different subsets of the population fall within the likely range of quarterly earnings (for example, TANF recipients versus leavers)? This involves knowing your target population's employment and earnings patterns and potentially using self-reported information at various points in time as a quality check on whether the wage records you've received from the UI agency look reasonable. For example, a TANF recipient will usually have lower employment outcomes than TANF leavers.

- Are there temporal or regional variations, including differences in rural and urban employment and seasonal variations in employment (for example, in tourist areas or during holiday shopping season) for localized analyses? This involves knowing the context of local labor markets—including whether your state has some large employers that dominate the labor market and whether the most common industries that operate in your state are subject to seasonal patterns.

Questions to consider while assessing data quality and completeness include:

- Do you feel confident that data problems have a small effect on the calculation of primary outcomes? Are they infrequent, small in magnitude, and unbiased?

- If problems occur, what can you reasonably do about them?

Common UI Wage Data Problems

A number of problems commonly encountered with UI wage data are listed below, followed by a list of useful checks that can help you investigate these problems. The GitHub repository includes R code that you can adapt to check for these issues:

- **THE DATA FILE IS MATCHED TO THE WRONG INDIVIDUALS.**

It is possible that identifiers were incorrect on a data request (for example, because of typos or data extraction errors), or that the agency's matching algorithm matched to individuals incorrectly (for example, through programming errors in the algorithm). The recent surge in "synthetic identities" that are the result of individuals using other people's social security numbers to secure employment or credit may also mean that some of the wage records in the file may be a reflection of identity theft or fraud rather than actual employment.[2] It may be difficult to identify which earnings records are mismatched, but a sign of a substantial problem is if some summary measures look like they are completely out of your range of expectations based on what you know about the TANF population you are studying.

- **THERE IS A PREVALENCE OF SHARED SOCIAL SECURITY NUMBERS.**

It is not uncommon for individuals, particularly in intergenerational families, to share SSNs for the purposes of work. If there is a higher than typical prevalence of SSNs in a particular part of the caseload, some individuals may appear as having extremely high earnings and working at multiple jobs simultaneously, while others who worked may not show up in the file if their SSNs in the TANF database are not correct.

- **THERE IS A PREVALENCE OF MULTIPLE OR CHANGING EMPLOYER IDS.**

It is also not uncommon for employers to have multiple IDs in the same quarter, or for an individual's earnings to be reported twice to the UI agency, such as from both a temp agency and from the referred employer. As an example, a large employer in southern California once reported its earnings under two different employer IDs. This made it hard initially to find the duplicate wage records. When these double-counted earnings were summed, earnings levels in the Los Angeles area looked like they were rising.

- **DATA HAVE BEEN RETRIEVED BEFORE EMPLOYERS HAVE REPORTED EARNINGS FOR SPECIFIC QUARTERS.**

Many states require employers to report quarterly earnings by the end of the month following the end of the quarter. In the TDI team's experience, the majority of earnings records are reported within six months of the close of the quarter. However, sometimes employers experience delays in reporting earnings to the UI agency. One large employer delay in a quarter can affect the estimates of employment rates of a caseload. It is important to have a data updating routine that incorporates overlapping quarters on

2 McClennan and Asaro (2017).

consecutive requests so that as new earnings records are added to those quarters, data can be reconciled. The most common update is a new earnings amount for a quarter that was not previously there (which might indicate late employer reporting). It is also common that earnings amounts may be updated over time. This means that an updated UI wage data transfer might show a different earnings amount from one employer for the same individual and quarter.

- **THERE IS INCOMPLETE DATA FOR THE MOST RECENT QUARTERS ON A FILE.**

The most recent quarter or two on a data extract may be incomplete because there is typically a lag between employers reporting wage data for the quarter and statewide earnings data being made available to the data analysts. The most recent quarter with complete data can be ascertained by looking at trends in percentage employed and earnings levels.

- **THERE ARE EXTREMELY HIGH OR LOW QUARTERLY EARNINGS PER PERSON OR PER INDIVIDUAL JOB.**

Outliers can appear as single records with an especially high or low earnings amount. Outlier values can also be found when individuals have multiple earnings records in the same quarter, with reasonable earnings amounts for each record but when summed appear too high. Determining whether certain earnings are "high" or "low" may require a larger discussion—potentially including case workers and program managers in the conversation.

Useful Checks for Identifying Problems in UI Wage Data

- **CHECK FOR MISSING OR INCOMPLETE DATA.**

The UI agency will return data that reflect quarterly earnings reported by employers, so there will not be records in the system for those who were never employed during the time frame covered by the data request. This means that missing data are sometimes correct but can also sometimes indicate a larger problem that may affect your analysis. Considerations include:

- whether the data you have are for the correct individuals, and whether you are missing data for individuals due to an identifier issue listed above
- whether the employment status and earnings levels are consistent with self-reported data
- whether missing data occur more often than expected in some quarters than in others, outside of typical economic trends such as an uptick in the fourth quarter as the demand for retail workers increases
- whether missing data occur systematically, such as by employer type or earnings amount

- Earnings for individuals may be missing but not because of an error: For example, your sample may have a large proportion of individuals who work out of state. Additionally, growth in the gig economy in recent years may mean that individuals' real earned income is not on the UI wage data file. Some states do not require employers to report quarterly wage records where earnings fall under a certain threshold (in Texas, for example, that threshold is $500), so individuals with low earnings in a particular quarter may not show up on the file.

- Missing data may indicate a data quality issue. For example, some employers may be delayed in reporting data for particular quarters, as described earlier.

To find and diagnose missing or incomplete data issues, you can calculate the "match rate" of the UI wage data file to the individuals in your sample by matching the identifier in the data file with the corresponding one in your request file. This match allows you to determine the percentage of individuals in your sample with a record in the UI wage data file. (If a SSN is in both files, use this to link the two files and estimate. Otherwise, use a common individual or group ID.) Use your team's knowledge of the groups of individuals in the data request to set your expectations for the match rate. It is useful to cross-check the information with self-reported employment and earnings data in corresponding time periods or participation status in employment placement services.

- Did you find individuals for whom you did not request data? If so, follow up with the data provider. This may signal a more general matching problem (for example, improper sorting and filtering of records).

- Is your match rate much lower or higher than expected? If so, you may have inaccurate identifiers.

■ Check for exact and partial duplicate records.

Data files should always be checked for exact duplicates in all fields. Normally, you would keep one record and drop the duplicate records. However, the presence of more than a handful of exact duplicate records in a source file could signal a more general matching problem. If possible, follow up with the data provider if you find an unexpectedly large number of exact duplicate records within a single source file.

UI wage data records should also be checked for partial duplicates, of which there are two types:

- There are two or more records that have **the same earnings amount, identifier(s), and quarter, but have different employer IDs.** You should investigate whether these are *real* duplicates (drop one or more records) or show earnings for different jobs (keep all records). Unusually high numbers of these types of partial duplicates may suggest that some large employers are transitioning to a different employer ID, rather than an individual who has exactly the same earnings amount from each employer.

- There are two or more records that have **the same employer, SSN, and quarter, but a different earnings amount.** You should investigate whether these types of partial duplicates represent two separate job spells within the same quarter (for an example, an employee of a large supermarket chain who works in two different stores during the same quarter), or whether only one of these records should be counted. If you are updating files (that is, new data for an overlapping quarter), the later record is likely the most accurate one.

It is often useful to identify individuals who have partial duplicates and then look at all of their records over time. Quarterly earnings histories by employer ID may give you a sense of whether multiple jobs were held or whether it is a changing employer ID issue.[3] If your file comes with an employer name, that may also help, though many employment agencies are reluctant to release this information. Note that employer IDs and names are very sensitive data and should be handled with great care, accessible only to staff members who actually need this information (for example, the analysts who are performing detailed data quality checks).

CHECK FOR OUT-OF-RANGE VALUES AND OUTLIERS.

UI wage data should be checked for out-of-range values to make sure that all fields have reasonable values. For example, if the quarter field is presented as a date, make sure that the date ranges for that field are all valid.

The earnings field should be checked for outliers (values that are a lot higher or lower than expected). The outlier definition will differ based on your expectations. Outliers could signal a problem with the UI wage data (for an example, an employer leaving out the decimal point when reporting someone's earnings) or a problem with that individual's SSN (which could mean that it was recorded incorrectly and belongs to someone outside of the sample with higher earnings, or that multiple people are working under an assumed SSN). Checks for outliers should be done at the beginning of the process when the file includes one record per person-quarter-employer (where each row represents one person in one quarter at one job), and then again once the earnings across employers in the same quarter are summed for each person. The latter check is important because sometimes you will first notice outlier amounts for an individual after earnings from each of the person's jobs are summed to a single quarterly total.

- A good first step is to look at the distribution of quarterly earnings at each job to determine whether they look reasonable (see Table 4, which shows an earnings distribution for a UI wage data file). Based on the distribution across dollar ranges in Table 4, your team might initially consider records with $10,000 or more of earnings

3 When possible, it is extremely useful to collect the employer's Federal Employer Identification Number (FEIN). For example, it is easier to perform data checks on partial duplicates as well as to do more data exploration with a FEIN than with a state's employer ID number. However, employer IDs are highly sensitive data and releasing FEINs may not be permissible. In such cases, it is worth asking whether the agency can at least provide unique "pseudo" (anonymized) FEINs that can be linked to other de-identified data sources.

Table 4. Distribution of Quarterly Earnings per Job

EARNINGS RANGE	NUMBER OF QUARTERLY EARNINGS RECORDS	PERCENTAGE OF TOTAL NUMBER OF EARNINGS RECORDS
$10 - $49.99	110	1.7
$50 - $99.99	199	3.0
$100 - $499.99	832	12.5
$500 - $999.99	731	11.0
$1,000 - $1,999.99	891	13.4
$2,000 - $4,999.99	2,358	35.5
$5,000 - $9,999.99	1,416	21.3
$10,000 - $19,999.99	78	1.2
More than $20,000	2	0.1

for one job in one quarter as possible outliers, but you would need to investigate specific cases more closely to be sure. You would also want to investigate some of the lower-end values.

- A useful follow-up check to do after restructuring the data to one record per individual is to look at the number of outliers by quarter. If it looks like outliers were all occurring in the same quarter, this could point to a problem with a particular file, or a systematic data extraction error for a particular quarter.[4]

CHECK FOR UNEXPECTED DATA CHANGES BETWEEN FILES.

As noted earlier, there may be some changes between files requested at different times due to UI wage data being updated by employers. When receiving multiple data shipments, it is important to request some overlapping data to examine the number of exact duplicates across shipments and the number of updates across shipments. For example, if you find that the overlapping quarters had no exact duplicates or updates, there may be an issue with the employer ID (for example, the agency may have sent a state employer ID on one file and a Federal Employer Identification Number on another). If you find that earnings between two file shipments changed for almost every quarter for individuals, it may be a programming problem (for example, one quarter of earnings may have been shifted to a different quarter). As a rule, when checking UI wage data, always look at trends by quarter in the number of people working, the average earnings, the maximum and minimum earnings, and the standard deviation. Look for any spikes or drops that can't be explained by seasonality.

4 See Hendra (2019).

- **CHECK FOR SHARED SSNS.**

As mentioned earlier, in some locations or populations there is sometimes evidence that people are sharing the same SSN for employment purposes. It is useful to check the data files to see if it seems like people have too many jobs (for example, 10 or more) in a quarter. These quarters will likely show outlier values of earnings, as well. Usually this is a rare event, but if it occurs frequently, it might be a good idea to check if it has affected results for specific subpopulations.

Resolving UI Wage Data Problems

In almost all situations, it is most straightforward to identify, follow up on, and resolve data issues if you have a direct data agreement for identifiable information with the state UI agency. For other types of data linkage arrangements, such as having more limited access to identifiable data or relying on a third party organization for data access, you may have fewer options at your disposal for examining the sources of data problems or for fully correcting for issues that you've identified. It may be useful to discuss acceptable thresholds of data quality that can still yield useful information for your purposes with your program evaluation team and develop standard rules for common data issues. For example, you might decide to take the higher earnings amount for any partial duplicate (same person, quarter, and employer, but different earnings amount), if it appears that either deleting a set or summing the amounts do not give you substantially different results.

The companion GitHub repository includes a folder with R Markdown files that demonstrate how to check for these common data problems. It also includes a summary table that serves as an at-a-glance guide on how to address the data problems described in this section. The synthetic data files used for these checks are also provided so that you can set up and run the code.

Data Processing and Restructuring

After you have checked your UI wage data file for data quality issues, you can proceed to data processing and restructuring data to produce a useful analysis file. Creating meaningful employment measures from state UI wage data involves restructuring the data to a file that sums earnings per quarter per employer (or per group of employers, or all employers). The restructured file can then be used to create relevant person-level employment-related outcomes on an analysis file. The GitHub repository includes another folder with SQL code that demonstrates how to link and restructure your UI wage data file for analysis. The synthetic data files used in this code are also provided so that you can set up and run the code.

The top table in Figure 3 shows a typical UI wage data file structure for one person that includes SSN, quarter of employment, employer ID, and earnings for one individual. Depending on the contract or request, the file may also include employer names and North American Industry Classification System (NAICS) industry codes. States with enhanced UI wage data availability may have information on weekly or monthly earnings or other job

Figure 3. Restructuring Raw UI Wage Data to a Person-Quarter-Level Data File for SSN 123456789

Data Structure for a Typical UI Wage Data Extract

SSN	QUARTER	EMPLOYER ID	EARNINGS
123456789	2007Q1	12	500
123456789	2007Q1	15	1000
123456789	2007Q2	20	2000
123456789	2007Q3	12	550
123456789	2007Q3	13	1000
123456789	2007Q4	10	100
123456789	2007Q4	20	2560
123456789	2008Q2	12	505
123456789	2008Q2	13	10

+

Information from TANF Cross-Reference File on SSN 123456789

SSN	PROGRAM START DATE	PROGRAM END DATE
123456789	7/15/2007	9/15/2010

Restructured Person-Quarter-Level Data File for SSN 123456789

SSN	PROGRAM START DATE	PROGRAM END DATE	QUARTER	QUARTER RELATIVE TO PROGRAM START DATE	NUMBER OF EMPLOYERS	TOTAL EARNINGS ($)
123456789	7/15/2007	9/15/2010	2007Q1	-2	2	1500
123456789	7/15/2007	9/15/2010	2007Q2	-1	1	2000
123456789	7/15/2007	9/15/2010	2007Q3	0	2	1550
123456789	7/15/2007	9/15/2010	2007Q4	1	2	2660
123456789	7/15/2007	9/15/2010	2008Q1	2	0	0
123456789	7/15/2007	9/15/2010	2008Q2	3	1	515

characteristics. Typically, a file will have multiple rows per person for each quarter that represent different jobs held in that quarter.

Restructuring Data

UI wage data are not ready for analysis in raw form. Note that there is more than one way to process and restructure the data to conduct your desired analysis. One straightforward way is shown in Figure 3, where the desired analysis file structure is a person-quarter–level file that includes an identifier, relevant information from the TANF cross-reference file (for example, program start and end dates), calendar quarter, number of employers, and earnings summed across employers. If you have research questions about employers, you can create additional columns on earnings from specific employers or groups of employers.

The example individual from the raw UI wage data extract would be restructured to six rows, which includes a row with no earnings in 2008 Quarter 1, as shown at the bottom of the figure.

The GitHub repository includes a [folder on creating outcomes]() with SQL code that demonstrates how to restructure these data for all employers as well as for specific employers.

Outcome Creation

Creating Intermediate Outcomes

The person-quarter–level file allows you to then create quarterly and annual outcomes as needed for analysis. One way to do this is to transform the data to a wide file with one row per person, with $0's imputed for any quarter that does not appear on the file for an individual. The assumption is that that individual was not employed for that quarter. Figure 4 shows the transformation of the individual's earnings data from Figure 3 transformed to one wide record, with a separate quarterly earnings amount for each quarter. You can create a corresponding "employed" flag for any quarter with a non-zero earnings amount. You can also create quarterly variables that are relative to a particular reference date depending on the analysis you are trying to do. A reference date might represent a date used for a cohort definition, relative to some event, or to the date of intake for a study. For example, if you are examining employment or earnings patterns for TANF leavers, you might create measures that reflect the number of quarters before or after individuals left the TANF rolls, which are called "relative measures," so that you can line up outcomes for people who left TANF at different points in time.

Note that for individuals whose identifiers were submitted to the UI agency and did not appear on the UI wage data file that the agency returned, you would assume that they were never employed in a UI-covered job for the quarters that you requested. In that case you would need to include those individuals on your individual-level data file and impute $0's for earnings in all quarters in the relevant follow-up period.

One advantage of using SQL code to transform these data is that you can easily create these measures selectively and as needed. The Github repository includes code to create quarterly calendar variables separately from code to create quarterly relative variables

Figure 4. Transformed Person-Level Data with Quarterly Outcomes

SSN	Reference Date	Earnings, 2007Q1	Earnings, 2007Q2	Earnings, 2007Q3	Earnings, 2007Q4	Earnings, 2008Q1	Earnings, 2008Q2
123456789	15JUL2007	$1,500.00	$2,000.00	$1,550.00	$2,660.00	$0.00	$515.00

Some examples of reference dates:

- Date of TANF enrollment
- Date of training enrollment
- Date of training completion
- Date of TANF exit

Employed, 2007Q1	Employed, 2007Q2	Employed, 2007Q3	Employed, 2007Q4	Employed, 2008Q1	Employed, 2008Q2
1	1	1	1	0	1

Employed, Pre-Q2	Employed, Pre-Q1	Employed, Relative Q0	Employed, Relative Q1	Employed, Relative Q2	Employed, Relative Q3
1	1	1	1	0	1

(using program start date as the reference date). You may also choose to combine these code segments to create both calendar and relative outcomes at once.

Creating standard employment-related measures

Researchers have developed some standard measures of employment-related outcomes that can be created from UI wage data. The following list includes adapted measures that MDRC developed for and used in the U.S. Employment Retention and Advancement study, a national evaluation of 16 demonstrations of welfare-to-work models deployed in the late 1990s and early 2000s:[5]

- employment (whether someone is working for pay) and frequency of employment
- employment stability (whether someone continues working, in either the same job or a different job)
- earnings (earned income from a job, which is affected by many things such as hours and job schedule)
- employment and earnings history measures (sums or averages in the year or two before a program, like TANF, or an intervention or milestone started). Measures of employment or earnings history should capture changes in trends such as the pre-program dip often observed in TANF populations (sometimes called the Ashenfelter Dip).[6]

5 See Hendra, et al. 2010.

6 Heckman and Smith (1995).

- advancement (improvements in the quality of work, including better schedules, higher pay, better positions, or more job benefits). Although many advancement measures cannot be directly measured using UI wage data, you can create measures based on earnings differences over time that may be indicators of advancement.

- sector-based employment (employment in a particular field of work)

Note that the examples of the measures presented below are not all-inclusive. The analysis file can be used to create outcomes for different time frames and follow-up periods, and even for descriptive “baseline” purposes. Some of the examples of measures presented in this section show how to create an employment-related measure after a specified reference date (which, as shown in Figure 4, can be a variety of events, activities, or milestones, such as enrollment in a program, completion of a program, or TANF exit). These measures can also be created in the same way for time periods *before* the reference date (for example, ever employed or total earnings in the year before enrolling in sector-specific training), or for shorter time frames (such as two or three quarters before or after an event, which are approximately six or nine months of pre- or post-follow-up). The main limitation is the inability to report on smaller-than-quarterly time frames, though some states (such as Illinois) are starting to report UI wage data each month.

The GitHub repository also includes SQL code that demonstrates how to create the annual measures of employment and earnings, employment spell measures, and employer-based measures described below. While the interim quarterly 0/1 measures are shown in the examples for clarity, the SQL code demonstrates how to efficiently create each of these outcomes directly from the person-quarter-level file.

- **EMPLOYMENT AND RETENTION MEASURES.**

“Employment retention” refers to the extent of an individual’s labor force participation and can be measured by calculating employment rates (by creating an indicator for ever employed during a follow-up period, as shown in Table 5), frequency of employment (by calculating the proportion of quarters an individual is employed during a follow-up period, as shown in Table 6), and stability of employment (by counting consecutive quarters employed or not employed, as in Tables 7 and 8 and Figure 5).

Table 5. Ever Employed

The percentage of participants who ever worked in a UI-covered job in the follow-up period

ID	EMPLOYED, Q1	EMPLOYED, Q2	EMPLOYED, Q3	EMPLOYED, Q4	EVER EMPLOYED, YEAR 1
0001	0	0	0	0	0
0002	1	0	0	0	1
0003	1	1	1	1	1

Table 6. Average Quarterly Employment

The frequency of employment in a UI-covered job in the follow-up period

ID	EMPLOYED, Q1	EMPLOYED, Q2	EMPLOYED, Q3	EMPLOYED, Q4	AVERAGE QUARTERLY EMPLOYMENT, YEAR 1
0001	0	0	0	0	0
0002	1	0	0	1	.5
0003	1	1	1	1	1

Table 7. Had an Employment Spell of At Least Four Quarters

Continuous employment retention is an employment stability measure.

ID	EMPLOYED, Q1	EMPLOYED, Q2	EMPLOYED, Q3	EMPLOYED, Q4	EMPLOYED, Q5	EMPLOYED, Q6	EMPLOYED, Q7	EMPLOYED, Q8	EMPLOYMENT SPELL OF AT LEAST FOUR QUARTERS
0001	0	0	0	0	0	0	1	1	0
0002	1	0	0	1	1	1	1	0	1
0003	1	1	1	1	1	0	0	1	1

NOTE: The yellow cells show employment spells of at least four quarters.

Figure 5 displays an example scenario where the person has high employment stability, even though job stability rates are much lower. High employment stability has proved to be more important than high job stability as an indicator of retention and potential advancement opportunities.

EARNINGS AND ADVANCEMENT MEASURES.

"Advancement" refers to improvements in fringe benefits received, working conditions, or opportunities for promotions. Although that information is not available in UI wage data, increases in earnings that are not fully explained by increases in employment retention could be used as a gross indicator of advancement. Increases in earnings provide a clear and direct measure of economic mobility. However, earnings can increase for reasons other than advancement. This is because earnings levels are directly impacted by employment rates. All else being equal, increases in any of the employment retention measures discussed in the previous section will increase earnings. This is because average earnings include zeros for people who were not working during the follow-up period. Consequently, increases in the proportion of the follow-up period with employment will directly translate into earnings increases. One indicator of advancement is seen when increases in earnings are larger, in percentage terms, than increases in employment retention. This is an effort to

Table 8. Length of Longest Employment or Unemployment Spell, in Quarters

Another employment stability measure

ID	EMPLOYED, Q1	EMPLOYED, Q2	EMPLOYED, Q3	EMPLOYED, Q4	EMPLOYED, Q5	EMPLOYED, Q6	EMPLOYED, Q7	EMPLOYED, Q8	LENGTH OF LONGEST EMPLOYMENT SPELL	LENGTH OF LONGEST UNEMPLOYMENT SPELL
0001	0	0	0	0	0	0	1	1	2	6
0002	1	0	0	1	1	1	1	0	4	2
0003	1	1	1	1	1	0	0	1	5	2

NOTE: The yellow cells show the longest employment cells for each person. The grey cells show the longest unemployment spell for each person. The blue columns show the calculated length, in quarters, of the longest employment and unemployment cells for each person in this example.

Figure 5. Percentage of Quarters Employed, All Employers or Individual Employer

Another employment stability and job stability measure

SSN	QUARTER	EMPID
123456789	2007Q1	12
123456789	2007Q1	15
123456789	2007Q2	20
123456789	2007Q3	12
123456789	2007Q3	13
123456789	2007Q4	10
123456789	2007Q4	20
123456789	2008Q2	12
123456789	2008Q2	13

SSN	QUARTERS EMPLOYED, ANY EMP (%)	QUARTERS EMPLOYED, EMPID 10 (%)	QUARTERS EMPLOYED, EMPID 12 (%)	QUARTERS EMPLOYED, EMPID 13 (%)	QUARTERS EMPLOYED, EMPID 15 (%)	QUARTERS EMPLOYED, EMPID 20 (%)
123456789	.83	.17	.5	.33	.17	.33

NOTE: In this example, percent values are calculated for each distinct employer identifier. For example, this individual has earnings from Employer 10 (in white) during one out of the six quarters beginning in 2007Q1 and ending in 2008Q2; from Employer 12 (pink) in three out of six quarters, and so on.

"tease out" the increases in earnings that are due to working more quarters from increases in earnings that are due to other factors (such as working for more hours, or at higher wages, or for more weeks). The next three tables describe different advancement measures created from quarterly earnings data.

Total annual earnings, as shown in Table 9, can be affected by numbers of quarters, weeks, or hours worked, and wage rates; increases may suggest possibility of advancement, but other measures need to be assessed.

Table 9. Total Annual Earnings

ID	Q1 EARNINGS ($)	Q2 EARNINGS ($)	Q3 EARNINGS ($)	Q4 EARNINGS ($)	TOTAL ANNUAL EARNINGS, YEAR 1 ($)
0001	0	0	0	0	0
0002	200	0	0	1,861	2,061
0003	1,250	4,250	1,250	1,300	8,050

Earnings of $3,500 or more corresponds to working full time at $7.50 an hour for all weeks in a quarter and may represent advancement for many TANF recipients. For states with higher minimum wages, a variable that represents a higher wage rate could be created (for example, earnings of $7,000 or more a quarter corresponds to $15 an hour). Table 10 shows how the percentage of quarters in a year with earnings of $3,500 or more can be constructed using quarterly earnings data.

Table 10. Percentage of Quarters with Earnings of $3,500 or More

ID	Q1 EARNINGS ($)	Q2 EARNINGS ($)	Q3 EARNINGS ($)	Q4 EARNINGS ($)	QUARTERS WITH EARNINGS ≥ $3,500, YEAR 1 (%)
0001	0	0	0	0	0.0
0002	200	0	0	1,861	0.0
0003	1,250	4,250	1,250	1,300	0.25

Table 11 presents an advancement measure that indicates whether an individual was earning more at the end of a follow-up period than at the start of it. Comparing the quarter of highest earnings at the beginning of the follow-up period with the quarter of highest

Table 11. Had Quarterly Earnings Increase by $250 or More in Three Years

ID	HIGHEST QUARTERLY EARNINGS, Y1 ($)	HIGHEST QUARTERLY EARNINGS, Y2 ($)	HIGHEST QUARTERLY EARNINGS. Y3 ($)	HIGHEST QUARTERLY EARNINGS, Y4 ($)	EARNINGS INCREASED BY $250 OR MORE Y4 VERSUS Y1
0001	0	0	500	0	0
0002	1,861	1,532	2,086	2,172	1
0003	4,250	1,350	1,450	2,000	0

earnings three years later can pick up on long-term advancement. This advancement measure reflects earnings growing over time.

- **SECTOR-SPECIFIC OUTCOMES.**

Since UI wage data often include the North American Industry Classification System (NAICS) code, the listed outcomes above also can be created separately for each sector of interest (see Figure 6). The synthetic dataset provided in the GitHub repository does not include a NAICS code, so the code for sector-specific outcomes is not provided. However, you can use the same code logic as provided for the employer-specific outcomes.

The measures described so far are at the individual level. It is often useful to produce measures at the firm or industry level as well.[7] Box 4 describes an approach to estimating job turnover; creating this and similar measures are covered in the Coleridge Institute's ADA training series.

Linking Individual-Level Wage Data to Case-Level TANF Data

Once you have created a person-level dataset with your desired employment-related outcomes, you can link these data back to your case numbers and then create a case-level dataset using the date-stamped information on your cross-reference file, as shown in Figure 7.[8]

Depending on your analysis question, you may choose to leave earnings data as missing for the child-only case (00006), rather than imputing 0s as shown in the example. You can then link this information to any other case-level TANF data to explore other relevant measures, such as income or benefit receipt information.

7 See Andersson, Holzer, and Lane (2005) for ideas and measures along these lines.

8 The code provided in the GitHub repository demonstrates how to link the TANF cross-reference file with the UI wage data file using a common identifier (SSN) but does not demonstrate conversions to case-level analysis.

Figure 6. Employed in a Specific Sector

NAICS codes can help to identify jobs in a certain sector. It is important to note that these are not occupation codes.

SSN	QUARTER	EMPID	NAICS	EARNINGS
123456789	2007Q1	12	236220:Construction	500
123456789	2007Q1	15	236220:Construction	1000
123456789	2007Q2	20	561110:AdminServc	2000
123456789	2007Q3	12	236220:Construction	550
123456789	2007Q3	13	452990:Retail	1000
123456789	2007Q4	10	452990:Retail	100
123456789	2007Q4	20	561110:AdminServc	2560

SSN	EMPLOYED IN CONSTRUCTION, 2007	EMPLOYED IN RETAIL, 2007	EMPLOYED IN HOME HEALTH CARE, 2007
123456789	1	1	0

NOTE: NAICS=North American Industrial Classification System.

Box 4. Creating Firm- or Industry-Level Measures

If you have employer or industry IDs on your Unemployment Insurance (UI) wage data file, it is often useful to create measures of firm or industry level wages and turnover both among TANF recipients and for the general state economy. Turnover captures the proportion of employees entering and leaving an employer or industry. A key measure of turnover is given by:

$$Turnover_{2018\ to\ 2019} = \left(\frac{[Joiners_{2018\ to\ 2019} + Leavers_{2018\ to\ 2019}\ ]}{Average\ Employees_{2018\ to\ 2019}}\right)$$

In and of itself, high turnover is not necessarily a bad thing. For example, high-growth firms can show high turnover. A more precise measure is called *churn*. One can calculate churn (the proportion of employees entering and leaving, adjusted for growth/decline) as:

$$Churn_{2018\ to\ 2019} = \left(\frac{[(Joiners_{2018\ to\ 2019} + Leavers_{2018\ to\ 2019}\ ) - \ |\Delta Total\ Employees_{2018\ to\ 2019}|]}{Average\ Employees_{2018\ to\ 2019}}\right)$$

Several other useful analyses can be done at the firm or industry level. This box just scratches the surface of what is possible.[a]

NOTE: [a]For another example of creating turnover measures at the firm level, see Cynthia Miller, Vanessa Martin, and Gayle Hamilton, *Findings for the Cleveland Achieve Model: Implementation and Early Impacts of an Employer-Based Approach to Encourage Employment Retention Among Low-Wage Workers* (New York: MDRC, 2008).

Figure 7. Linking UI Wage Data File with TANF Cross-Reference File to Conduct Case-Level Analysis

Returned UI Wage Data File, Checked and Processed

SSN	TOTAL EARNINGS, 2018	TOTAL EARNINGS, 2019
111111111	0	0
222222222	5324	2261
333333333	10500	4300
444444444	200	0
555555555	0	1500
666666666	0	0

UI Wage Data Linked to Case IDs

CASE_ID_2018	CASE_ID_2019	SSN	TOTAL EARNINGS, 2018	TOTAL EARNINGS, 2019
00001	00001	111111111	0	0
00002	00001	222222222	5324	2261
00003	00003	333333333	10500	4300
00003	00003	444444444	200	0
00004	00004	555555555	0	1500
00005	00005	666666666	0	0
00006	00006			

Case-Level UI Wage Data

CASE ID	TOTAL HOUSEHOLD EARNINGS, 2018	TOTAL HOUSEHOLD EARNINGS, 2019	ANY ADULT EMPLOYED, 2018	ANY ADULT EMPLOYED, 2019	NUMBER OF ADULTS EMPLOYED, 2018	NUMBER OF ADULTS EMPLOYED, 2019
00001	0	2261	0	1	0	1
00002	5324	MISSING	1	MISSING	1	MISSING
00003	10700	4300	1	1	2	1
00004	0	1500	0	1	0	1
00005	0	0	0	0	0	0
00006	0	0	0	0	0	0

NOTE: The row highlighted in blue represents a decision to impute 0 earnings for the child-only TANF case.

Considerations for Data Analysis and Interpretation

Given some of the limitations of UI wage data listed earlier, agencies need to consider whether or to what extent UI wage data can meet their analysis needs. UI wage data can answer questions about historical trends for UI-covered jobs but may not be as useful in reporting out metrics on detailed job characteristics. Not fully understanding the implications of using UI wage data can potentially lead to misinterpreting employment-related findings from linking TANF and UI wage data. Examples of common issues to consider include:

- The completeness of UI wage data for individuals is often highly dependent on the accuracy and completeness of SSNs for each family member in a caseload. Populations with a higher prevalence of shared SSNs among family members or dummy SSNs (often used for children, although children are unlikely to be included in an employment analysis) may yield invalid results.

- Employment in non-UI-covered jobs may have implications for relevance in assessing program efficacy. For example, the noncoverage issue may vary in importance and relevance across different sectors.

- Hourly wages or month-to-month employment stability may or may not be necessary to address your program questions, given that UI wage data are quarterly and hours worked may not be accessible. A common error is to misinterpret increases in UI earnings as reflecting increases in hourly wages. While hourly wages could drive increases in total UI-covered earnings, earnings may also be higher due to increased weeks or hours worked. A growing number of states are able to provide information on hours worked. For example, Washington has had reported hours for the past 30 years, and the data are structured the same way as the quarterly data. There are also discussions of including data on job title. At least one state (Illinois) is now providing monthly UI wage data.

- UI wage data often lag by around four to six months, so consider how quickly you need the information.

- Your research questions will determine the most appropriate individuals to link to UI wage data and whether to estimate household-level employment measures. It is important to think through the analytical implications for whether to follow individual- or household/case-level employment-related outcomes, and what to do for child-only cases or households with more than one adult.

CONCLUSION

The aim of this toolkit is to help TANF agencies in your efforts to access, link, and prepare to analyze UI wage data. Although the toolkit covers many areas and scenarios, which can make the process sound overwhelming, one final note of encouragement: Many research teams have successfully collected and analyzed UI wage data over the years. The legal, technological, and policy climate is always changing, but the field is nearing 50 years of experience successfully using these data, and the expectation is that UI wage data will become even more useful and perhaps easier to securely obtain in the future. In the meantime, your questions, corrections, and comments are very much welcome. Please connect with the team at cdicode@mdrc.org.

APPENDIX
FOUNDATIONAL RESOURCES

The following resources include foundational insights about data sharing and integration, governance, legal agreements, ways to build agency capacity, and how to center racial equity.

Introduction to Data Sharing and Integration

Confidentiality Toolkit

Case Study Report - Iowa's Integrated Data System for Decision-Making (I2D2)

IDS Governance: Setting Up for Ethical and Effective Use

Legal Issues for IDS Use: Finding a Way Forward

Finding a Way Forward: How to Create a Strong Legal Framework for Data Integration

Writing Guide for a Memorandum of Understanding (MOU)

Sample Memorandum of Understanding and Data Sharing Agreement from the New York State Department of Labor

Coleridge Initiative Applied Data Analytics Courses

Technology for Civic Data Integration

Collaborative Institutional Training Initiative's (CITI) module on Ethical and Appropriate Uses of Administrative Data for Research and Evaluation

A Toolkit for Centering Racial Equity Throughout Data Integration

AISP Working Paper: Addressing Racial and Ethnic Inequities in Human Service Provision

Ideas for Centering Racial Equity in TDC Pilot Projects

References

Abraham, Katherine, John C. Haltiwanger, Kristin Sandusky, and James Spletzer. 2013. "Exploring Differences in Employment Between Household and Establishment Data." *Journal of Labor Economics* 31, S1: 129-172.

Andersson, Fredrik, Harry J. Holzer, and Julia I. Lane. 2005. *Moving Up Or Moving On: Who Gets Ahead in the Low-Wage Labor Market?* New York: Russell Sage Foundation.

Barnow, Burt S., and David Greenberg. 2015. "Do Estimated Impacts on Earnings Depend on the Source of the Data Used to Measure Them? Evidence from Previous Social Experiments." *Evaluation Review* 39, 2: 179-228.

Center on Budget and Policy Priorities. 2022. "Policy Basics: Temporary Assistance for Needy Families." Website: https://www.cbpp.org/research/family-income-support/temporary-assistance-for-needy-families.

Chase-Lansdale, Lindsay and Jeanne Brooks-Gunn. 2014. "Two-Generation Programs in the Twenty-First Century. *Future Child* 24(1): 13-39.

Dorsett, Richard, Richard Hendra, and Philip K. Robins. 2018. "Using Administrative Data to Explore the Effect of Survey Nonresponse in the UK Employment Retention and Advancement Demonstration." *Evaluation Review* 42, 5-6: 491-514.

Eberts, Randall W. 2015. "Toward a More Intelligent Workforce Development System." Pages 383–410 in Carl Van Horn, Tammy Edwards, and Todd Greene (eds.), *Transforming U.S. Workforce Development Policies for the 21st Century.* Atlanta: Federal Reserve Bank of Atlanta.

Evermore, Michele. 2020. "Long Lines for Unemployment: How Did We Get Here and What Do We Do Now?" (April). New York: National Employment Law Project.

Goerge, Robert M., Emily R. Wiegand, and Leah Gjertson. 2021. "Unpacking Data Use In State TANF Agencies." OPRE Report 2021-94. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Hawn Nelson, Amy, Della Jenkins, Sharon Zanti, Matthew Katz, T.C. Burnett, Dennis Culhane, and Katie Barghaus. 2020. *Introduction to Data Sharing and Integration*. Philadelphia: Actionable Intelligence for Social Policy at the University of Pennsylvania. Website: https://aisp.upenn.edu/resource-article/introduction-to-data-sharing-and-integration/.

Hayes-Greene, Deena, and Bayard P. Love. 2018. "The Groundwater Approach: Building a Practical Understanding of Structural Racism." The Racial Equity Institute. Website: https://www.groundwaterinstitute.com/groundwaterapproach.

Heckman, James J., and Jeffrey A. Smith. 1995. "Ashenfelter's Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies." Department of Economics Research Reports, 9505. London, ON: Department of Economics, University of Western Ontario.

Hendra, Richard, Keri-Nicole Dillman, Gayle Hamilton, Erika Lundquist, Karin Martinson, and Melissa Wavelet. 2010. *How Effective Are Different Approaches Aiming to Increase Employment Retention and Advancement? Final Impacts for Twelve Models*. New York: MDRC.

Hendra, Richard. 2019. "TDC Quick Tips: Dealing With Outliers in Human Services Data." *TANF Data Collaborative* (blog), June. https://tanfdata.org/content/tdc-quick-tips-dealing-with-outliers-in-human-services-data?category%5B%5D=3.

McClennan, Mark and Jennifer Asaro. 2017. "ID Analytics Finds Problem of Synthetic Identities is on the Rise." *Business Wire*, July. Website: https://www.businesswire.com/news/home/20170726005310/en/ID-Analytics-Finds-Problem-of-Synthetic-Identities-is-on-the-Rise.

McDaniel, Marla, Tyler Woods, Eleanor Pratt, and Margaret C. Simms. 2017. "Identifying Racial and Ethnic Disparities in Human Services: A Conceptual Framework and Literature Review." OPRE Report 2017-69. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Miller, Cynthia, Vanessa Martin, and Gayle Hamilton. 2008. *Findings for the Cleveland Achieve Model: Implementation and Early Impacts of an Employer-Based Approach to Encourage Employment Retention Among Low-Wage Workers*. New York: MDRC.

Office of Child Support and Enforcement. 2021. *A Guide to the National Directory of New Hires*. Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services. Website: https://www.acf.hhs.gov/css/training-technical-assistance/guide-national-directory-new-hires.

Randall, Sean, Adrian P. Brown, Anna M. Ferrante, and James H. Boyd. 2019. "Privacy Preserving Linkage Using Multiple Dynamic Match Keys." *International Journal of Population Data Science* 4, 1.

Safawi, Ali and LaDonna Pavetii. 2020. "Most Partnets Leaving TANF Work, But in Low-Paying, Unstable Jobs, Recent Studies Find." Washington, DC: Center on Budget and Policy Priorities. Website: https://www.cbpp.org/research/family-income-support/most-parents-leaving-tanf-work-but-in-low-paying-unstable-jobs.

Wiegand, Emily, Robert Goerge, and Leah Gjertson. 2017. "Family Self-Sufficiency Data Center Creating a Data Model to Analyze TANF Caseloads." Chicago: Family Self-Sufficiency Data Center. Website: https://www.chapinhall.org/wp-content/uploads/IB_FSSDC_082917.pdf.

Yang, Edith, Aurelia De La Rosa Aceves, and Gloria Tomlinson. 2019. *How Community-Based Organizations Can Use New York State Employment and Wage Data: Learning from the New York City Demonstration (2016-2018)*. New York: MDRC.

Yang, Edith and Richard Hendra. 2018. "The Importance of Using Multiple Data Sources in Policy Assessments: Lessons From Two Conditional Cash Transfer Programs in New York City." *Evaluation Review* 42, 5-6: 550-574.

ANNOTATED BIBLIOGRAPHY
ADDITIONAL TANF AND UI WAGE DATA PUBLICATIONS

Guidance for Wage Record Linkage and Use

This section contains resources for linking wage records, including technical and legal guidance, overviews of wage data sources, and general integrated data systems (IDS) and data linkage materials. Some of the resources are specifically intended for the wage and education data linkage audience.

A Guide to the National Directory of New Hires

This guide details who can access NDNH data and how. The NDNH, originally created for the purpose of child support enforcement, is a national database of wage and employment information (new hire, quarterly wage, and UI). NDNH data are only available to authorized agencies. State TANF agencies can gain access to carry out state responsibilities under programs funded under Part A of Title IV of the Social Security Administration. Other state agencies can gain similar access to carry out state duties.

Establishing a Standard Data Model for Large-Scale IDS Use

This expert panel report from Actionable Intelligence for Social Policy (AISP) provides high-level guidance for selecting data for an IDS and an overview of common types of data shared, including workforce, UI, and TANF, and discusses basic data standards—person, encounter, place, and time.

Joint Guidance on Data Matching to Facilitate WIOA Performance Reporting and Evaluation

From the federal Department of Labor and Department of Education, this resource provides information for states on requirements, procedures, and options for matching confidential Unemployment Compensation (UC) data from wage records with personal information from Vocational Rehabilitation records and personally identifiable info (PII) from education records. It was mainly created to help states with Workforce Innovation and Opportunity Act

(WIOA) reporting requirements involving data from quarterly wage records and education, and includes seven different methods of linking wage and education data.

Legal Issues for IDS Use: Finding a Way Forward

This report from AISP provides an overview of the common legal concerns and barriers for IDS, the foundational agreements (memos of understanding, data use licenses), relevant laws, and appendix examples. This is not specific to TANF and UI wage data but it is generally helpful for anyone starting to address legal concerns. It is mostly for people at the beginning of IDS construction.

Sources and Linking Strategies for Employment Data

This brief on the Statewide Longitudinal Data System (SLDS) offers a high-level overview of types of wage data and their key features. It is specifically aimed at users who are connecting education and employment data for SLDS, but also briefly mentions the utility of using TANF or other programs to show connections in outcomes.

State Wage Interchange System: Better Data for Stronger Workforce Programs

The SWIS system allows for assessing and reporting on state and local performance for workforce training and education programs. This report lists data available under SWIS and compares that with the Wage Record Interchange System (WRIS) and WRIS2 (SWIS expands on the WRIS systems).

Using UI Wage Data to Improve Program Employment Outcomes

This resource provides guidance for postsecondary education institutions that want to access UI wage data to analyze employment and labor market outcomes for their students and graduates.

Making Wage Data Work: Creating a Federal Resource for Evidence and Transparency

This brief from the National Skills Coalition provides an overview of sources of federal wage data, including a very helpful table that documents what each source covers as well as recommendations for action at both the state and federal levels to improve the data quality and completeness in each source.

Compendium of Administrative Data Sources for Self-Sufficiency Research

This joint report by MDRC and the federal Office of Planning, Research and Evaluation (OPRE) includes an overview of sources of self-sufficiency data, including TANF and UI records. It is a key resource to draw on for organizations interested in accessing data to measure long-term outcomes on economic mobility, public assistance receipt, and health and well-being.

How Community-Based Organizations Can Use New York State Employment and Wage Data

This MDRC publication is a New York State–specific guide for municipalities and community organizations that are considering requesting access to state UI wage data. It includes a discussion of key challenges, opportunities, and lessons learned related to accessing UI wage data for program participants and details the legal process between all parties involved in data sharing.

Roadmap for K-12 and Workforce Data Linkages

This publication from the Data Quality Campaign (DQC) offers examples of questions that can be answered by linking education and workforce data and includes tips for data matching and sharing. It has some overlap with AISP's guidance on linking data, such as how to create a shared vision as a state, engage working partners, and develop strong data governance processes.

Investigating Alternative Sources of Quarterly Wage Data

This report, prepared by the Urban Institute, discusses sources of wage data other than UI and how to access them, relevant rules and policies, and how the data are used in research (as of 2012). Sources include: NDNH, Census Longitudinal Employer Householder Dynamics, WRIS, and the Administrative Data Research and Evaluation project.

Data on Earnings: A Review of Resources for Research

This report from Mathematica describes the big categories of wage-related data sources. The primary focus is on administrative data, but the report also discusses customer program participant surveys and existing population surveys and includes numerous tables regarding different types of administrative data on earnings. While this report does not discuss how to access data, it includes considerations and recommendations for working with administrative data.

Enhancing Unemployment Insurance Wage Records: Potential Benefits, Barriers Opportunities

This is a 2014 report from a federal work group that surveyed state agencies, organizations that use UI data, and software companies that report wage records for employers. The report provides an overview of the state of UI data collection, use, and challenges.

Developing the Basis for Secure and Accessible Data for High Impact Program Management, Policy Development, and Scholarship

This edition of *The Annals* synthesizes cross-disciplinary perspectives on data infrastructures that need to be developed so that data providers and researchers can address national policy questions. The book is structured around three topics: privacy and confidentiality, data providers, and comprehensive strategies.

[Engagement with State and Local Government]

This handbook from the Center for Open Science provides guidance for state and local government agency leaders who are providing data and for internal or external organizations who are requesting data for analytics projects. It was created in response to calls from academic and research center–based investigators for the expanded use of government administrative data for evidence-building.

[Increasing Equity and Improving Measurement in the U.S. Unemployment System: 10 Key Insights from the COVID-19 Pandemic]

This report from the California Policy Lab and the California Employment Development Department highlights six insights about equity in access to UI benefits during the COVID-19 pandemic and four insights about measurement issues for these data.

TANF and Wage Record Studies and Reports

This section includes studies that have been done using TANF and wage data and reports that are specific to TANF and wage data linkage.

[TANF and Related Administrative Data Project]

This report illustrates work in Connecticut, Indiana, South Carolina, and Wisconsin enabled by grants from the Administration for Children and Families (ACF) for enhancing the use of TANF and related administrative data.

[Do State-Customized TANF Work Policies Actually Reduce Unemployment?]

This paper in *Social Science Quarterly* provides an example of how TANF and workforce data can be used for analysis. A difference-in-difference analysis was used to show that states implementing worker supplement programs achieve lower unemployment among women with low incomes.

[Families on TANF in Illinois: Employment Assets and Liabilities]

This 2003 study examined characteristics, circumstances, and job readiness of single-parent TANF cases in Illinois. The researchers used wage and UI data to show work patterns and earnings rates. of TANF and former TANF recipients.

[Measuring Employment Outcomes in TANF]

This 2018 Urban Institute report addresses the challenges of and opportunities for measuring employment outcomes in TANF. The report fleshes out issues primarily from the TANF agency perspective. The section on "Issues with Federal Sources of Employment Data" is particularly relevant for TANF and UI audiences.

[Testing Two Subsidized Employment Models for TANF Recipients]

This MDRC evaluation of two subsidized employment models for TANF recipients in Los Angeles uses both administrative data (earnings, TANF receipt, food stamps receipt, subsidized employment payroll records) and survey data.

[Use of Unemployment Insurance and Public Employment Services after Leaving Welfare]

This working paper from the W.E. Upjohn Institute for Employment Research looks at several outcomes for adults in households that recently received TANF benefits, including rates of joblessness, applications for and receipt of UI benefits, and participation in publicly funded employment services. The paper also examines the correlation between UI and employment services receipt and the return to work and independence from TANF.

[A Compilation of Select Papers from the 2003 Biennial National Research Conference]

One paper from this compilation, "Former TANF Recipients' Monetary Eligibility for Unemployment Insurance Benefits," demonstrates the types of questions that can be explored when linking TANF and employment data. The study highlight focuses on Wisconsin and matches data from TANF, workforce, food stamps, medical assistance, childcare, and the Wisconsin Work program with UI data.

TANF-Specific Guidelines or Publications

This section includes guidance, reports, and publications specifically related to TANF data.

[Welfare Rules Databook 2018]

The Urban Institute and OPRE compiled TANF policies as of July 2018 for all 50 U.S. states and the District of Columbia. Numerous tables show how states determine eligibility, compute benefits, what their work requirements are, and time limits on benefits. Each table lists all the states and whether or not they have certain policies. This resource is heavy on policy and lighter on data but it could be helpful for comparing policies across states and for understanding how that might affect data linkage.

A Bibliography of Studies Using TANF Linked Administrative Data

This is an Urban Institute overview of studies that used administrative data linkages to answer TANF research questions.

Confidentiality Toolkit

Chapter 3 of this toolkit produced by the Administration for Children and Families (ACF) Interoperability Initiative offers a high-level discussion about the case for sharing TANF data, relevant federal legislation and regulations, and guidance for sharing TANF data.

Family Self-Sufficiency Data Center: Needs Assessment Report

This report from Chapin Hall details results from a 2014 needs assessment of the Family Self-Sufficiency Data Center (FSSDC). Interviews and focus groups were conducted with researchers, policymakers, and administrators to understand more about how the FSSDC could support their needs around family self-sufficiency and well-being data.

Family Self Sufficiency Data Center: Creating a Data Model to Analyze TANF Caseloads

This Mathematica publication presents high-level guidance for agencies with TANF data access on how to restructure those data sets in order to answer questions about program management, evaluation, caseloads, etcetera. It recommends starting with a simple, small area of analysis and then expanding to larger projects once you have achieved success.

Measuring What Matters: Shifting TANF to an Outcome-Based Model

This brief from the National Skills Coalition presents the rationale for aligning TANF outcomes with WIOA common measures such as employment in the short term and in the long term, earnings level, and credential attainment. It could be used for making the case that it is important to be able to link DOL data with TANF data.

Tracking the Outcomes of Welfare Reform in Florida for Three Groups of People

This 2000 study used data from TANF, the state wage system, and telephone surveys to report on outcomes for Florida's TANF "leavers" (that is, former TANF recipients who left the program), "diverts" (individuals who started applying to TANF but did not complete the process), and "opt-outs" (Medicaid recipients who met TANF income qualifications but never applied). The study demonstrates basic questions that can be answered with TANF and wage data.

Understanding the Food Stamp Program Participation Decisions of TANF Leavers

In this paper, researchers matched administrative data from TANF with administrative data from the Food Stamp Program (FSP), now called the Supplemental Nutrition Assistance Program (SNAP). The paper also describes the record linkage process.

Examples of Wage- or Employment-Specific Studies

This section includes academic publications and research studies using wage and employment data.

Do Estimated Impacts on Earnings Depend on the Source of the Data Used to Measure Them? Evidence from Previous Social Experiments

This study examines the differences between using survey data and administrative data on earnings for impact estimates and emphasizes that there are important tradeoffs to consider when deciding what to use.

Measuring Employment and Income for Low-Income Populations with Administrative and Survey Data

This report from the Institute for Research on Poverty discusses the strengths and weaknesses of various income and employment data sources, including surveys, UI, and tax returns.

Self-Sufficiency of Former Foster Youth in Wisconsin

This report from the Institute for Research on Poverty used administrative data rather than interviews to understand employment, earnings, and public assistance receipt among former foster youth in Wisconsin.

The Importance of Using Multiple Data Sources in Policy Assessments: Lessons from Two Conditional Cash Transfer Programs in New York City

This MDRC study compared two cash transfer programs using both administrative data and survey data. Survey data showed positive employment effects for both programs, while the administrative data showed no statistically significant employment effects for either program. The results suggest that there are challenges with sample attrition in both survey and administrative data, and it may be necessary to triangulate data sources to depict program effects more accurately.

The Use of Linked Employer-Employee UI Wage Data

This 1997 paper from The University of Texas (UT) at Austin discusses the pros and cons of using linked employer-employee UI wage data and provides an overview of the standard wage data elements. The paper also provides succinct overviews of additional UT projects that used administrative data, including a welfare-to-work study led by Jacob France Center at the University of Baltimore.

Transforming U.S. Workforce Development Policies for the 21st Century

This compilation of papers on the changing landscape of workforce policy and programs, published by the W.E. Upjohn Institute for Employment Research, includes case studies, critical analyses, and recommendations for the future of this field.

Use of SNAP Benefits by UI Applicants in Michigan During the Great Recession

This paper from the W.E. Upjohn Institute for Employment Research examines SNAP use before and after UI application by linking and analyzing Michigan SNAP and UI data to understand trends during the recent Great Recession. This work could serve as an example for linking and analyzing UI data to understand the economic fallout from COVID-19.

A Methodology for Statistical Adjustment Under the Workforce Innovation and Opportunity Act

This is an “in the weeds” resource on WIOA that discusses DOL’s previous methods for setting performance targets for states, proposes an updated methodology for setting performance targets, and presents statistical results using the proposed model with current and simulated data.

The Washington State Merged Longitudinal Administrative Database

This paper provides an overview of how Washington State has used UI data to understand the effects of policy on employment. The authors augment UI data with voter, licensing, social service, income transfer, and vital statistics records to show earnings and employment trends by race and gender and to understand the impact of non-earnings income. Linking all these data sets allows for a comprehensive database of nearly all state residents. The paper also describes the data system and linkage processes.

Moving Up or Moving On: Who Gets Ahead in the Low-Wage Labor Market?

This book, published by the Russell Sage Foundation, examines the traits of employees with lower wages who experience upward mobility in the labor market, as well as traits of their employers. For TANF and UI audiences, Chapters 2 and 4 are particularly relevant. Chapter 2 describes the nature of the Longitudinal Employer Household Dynamics data and longitudinal employment measures. Chapter 4 demonstrates ways low-wage workers can transition into higher-paying employment.

[Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Reports from Employers Agree with Surveys of Individuals?]

This 1999 article in the *Journal of Labor Economics* explores the extent to which wage records reported by employers to state UI agencies can serve as a valid alternative to more costly retrospective sample surveys when measuring the impacts of employment and training programs for low-income people. The authors analyzed UI data and survey data from adults and youth with low incomes participating in the National Job Training Partnership Act study and found that impact estimates based on UI data and survey data were usually comparable. However, average earnings reported in surveys were higher than earnings reported in UI.

NEW YORK
200 Vesey Street
23rd Flr.
New York, NY 10281
Tel: 212 532 3200

OAKLAND
475 14th Street
Suite 750
Oakland, CA 94612
Tel: 510 663 6372

WASHINGTON, DC
750 17th Street, NW
Suite 501
Washington, DC 20006

LOS ANGELES
11965 Venice Boulevard,
Suite 402
Los Angeles, CA 90066

CHAPIN HALL AT THE UNIVERSITY OF CHICAGO

mdrc BUILDING KNOWLEDGE TO IMPROVE SOCIAL POLICY

COLERIDGE INITIATIVE

OPRE

ADMINISTRATION FOR CHILDREN & FAMILIES
Office of Family Assistance

AISP