# Creating a Longitudinal TANF Funding & Expenditures Dataset: Transformation Description

*Last updated Dec 23, 2024*

## Introduction

The Office of Family Assistance (OFA) collects and releases data on the federal funds awarded to and expended by states for the Temporary Assistance for Needy Families (TANF) program quarterly. Annually, OFA compiles this data and publicly releases a dataset summarizing TANF funding and expenditures. TANF has existed since 1996 and all states had to adopt the program by July 1997. The data that OFA release thus provide a rich longitudinal dataset for public servants, citizens, and academics interested in the evolution of the program. However, users who wished to use the data have historically had to do a substantial amount of work to understand the data and to compile it over time.

When these data accessibility issues were resurfaced by a recent Presidential Management Fellows (PMF) Challenge Team report, OFA reached out the ACF Tech Data Surge Team (henceforth, the Surge Team) for support. The PMF report made three recommendations:

1. That OFA take additional steps to showcase the TANF data, by creating blog posts and fact sheets
2. That OFA consolidate the data, by creating a single file which contains data for all years from 1997-2023 (present) and by making data accessible via dashboards
3. That OFA clarify the data by publishing descriptions of each data source and metadata files for each dataset.

OFA asked the Surge Team to focus on the final two recommendations. Specifically, OFA wanted the Surge Team to append TANF caseload and expenditure data for the full range of years available, to create the metadata files for these new datasets, and to develop a dashboard to showcase this data in Tableau. The present document outlines the steps taken to append the caseload data, append the expenditure data and define Tableau related variables (where applicable) for each dataset.

## Expenditure (Financial?) Data

### Description of the Data

The TANF expenditure data are currently published in one or more Microsoft excel workbooks for each fiscal year. In 2015, a revised version of the [ACF-196](https://www.acf.hhs.gov/sites/default/files/documents/ofa/instruction_completion_acf_196.pdf) instructions ([ACF-196R](https://www.acf.hhs.gov/sites/default/files/documents/ofa/acf_196r_instructions_final.pdf)) was released to help improve accounting:

Prior to FY 2015, TANF financial reporting expenditure categories did not fully reflect the wide range of benefits and services funded by federal TANF and state MOE dollars, causing states to categorize many activities simply as “other” and allowing certain activities to fall into multiple categories at once. This created confusion and inconsistencies that made analyzing spending information and comparing data across states problematic. Additionally, it was difficult to understand exactly how much money had been spent in a given fiscal year, due to the cumulative reporting nature of the previous accounting method.

In an effort to increase transparency and accuracy of the TANF financial data and eliminate ambiguities and inconsistencies without placing undue burden on states, OFA created the revised reporting form ACF-196R, which does two things:

The ACF-196R modifies and expands the list of expenditure categories and accompanying definitions (see Figure 4). It includes new categories such as child welfare, services to children and youth, and pre-kindergarten/Head Start. It also requires narrative descriptions of expenditures reported as “Other,” and “Authorized Solely Under Prior Law.”

The ACF-196R changes the accounting method to require states to report actual expenditures made in a fiscal year and make any subsequent revisions or corrections to the report for the fiscal year in which that expenditure occurred.[[1]](#footnote-2)

This also had the effect of changing the content of the workbooks resulting in two “stable periods” (periods during which the information collected in form ACF-196 does not change): 1997-2014, and 2015-2023 (present). In addition to the fundamental changes in the contents of the files, file formats change over time. There are three distinct periods for this: 1997-2009, 2010-2014, and 2015-2023 (present). There is also intra-period variation in file structure, including workbook names, worksheet names, and column names.

### File Preparation

Some files required manual preparation. For example, in the following years Line 6l (Non-Assistance Authorized Solely Under Prior Law) was labeled 6i: 1997, 1998, 1999, 2000, 2001, 2002, 2004, 2005, and 2009. We renamed many files in the period 2010-2023 to simplify their ingestion into Python. We also renamed a 2010 workbook sheet to align its name with others in the period, changing it from “Total State Expenditures” to “Total State Expenditure Summary”.

### Transformations

The complexity of the data demands addressing unique challenges within each period. This section is separated into subsections that identify universal transformations, as well as those applied only within a specific period.

#### Universal

As noted in the Description of the Data section, column names varied across files both within stable periods and within periods where file formats were consistent. To circumvent this problem, we mapped the column names in the workbooks to their corresponding line numbers. This ensured that any variation in naming conventions between the workbooks did not impact our ability to append data within stable periods.

In rare cases values in a worksheet were missing. We assume that a missing value means that a state made no expenditures in that category. Therefore, we set all missing values to 0. A similar approach is taken when summing across the federal and state levels to generate the total level. Some categories which exist at the federal level do not exist at the state level. To prevent missing values, we set values in these non-existent columns to 0 during this addition step. Two important distinctions between the zeroing in the addition step and appending steps should be noted: 1) a missing value still appears as missing at the appropriate level after zeroing during the addition step, and 2) if both the state and federal level are missing a value (which occurs often in earlier years, since a category may not have existed at all) the value appears as missing at the total level.

Some missing values appear as strings in the data which often results in numeric columns being read as string. After dealing with missing values (e.g. replacing “-“ with 0), we convert all columns to numeric. During this conversion we round values to the nearest integer.

In all years and across all tabs, the data is limited to the 50 states, D.C., and a U.S. Total Row. U.S. territories and Puerto Rico are not included in the data.

#### 1997-2009

State-level results are the sum of expenditures using State Family Assistance Grant (SFAG) Funds and expenditures on Assistance using Maintenance of Effort Funds in Separate State Programs (SSP). Beginning in 2010, this sum is precalculated in the public-facing workbooks, but in 1997-2009 this is not the case. Therefore, to generate the state tab in this period we explicitly sum the SFAG and SSP funds.

Similarly, from 1997-2009 carryover—the funds in a state’s budget that are residuals from prior years—is not included in the workbooks. To get this figure, we sum the federal unliquidated obligations (Line 9, 196) and unobligated balance (Line 10, 196) in the year prior to the current year. For example, to calculate carryover in 2009, we sum 2008’s federal unliquidated obligations and unobligated balance. This means that in 1997, the first year for which we have data, carryover is set to 0.

Across most worksheets, a “U.S. TOTAL” row is included which sums values across all states within that year. In years in which this row doesn’t appear, we create it by summing the values across all states.

#### 2010-2014

No period-specific changes were implemented.

#### 2015-2023

No period-specific changes were implemented.

### Mapping across the disjoint

As noted earlier, in 2015 the instructions for completing form ACF-196 were revised. This resulted in an increase in the number of reporting categories. Practically, this means that some lines in ACF-196R map directly to lines in ACF-196, some lines have no ACF-196 analogue, and in some cases multiple lines in ACF-196 need to be summed to arrive at a single line in ACF-196R. We created a crosswalk which maps ACF-196 columns to their corresponding ACF-196R counterparts. Then, after labeling column names with their respective line numbers for all files, we leveraged the crosswalk to perform the necessary renaming and summation to convert ACF-196 lines into ACF-196R lines. This mapping is imperfect. For example, prior to 2015 (ACF-196) college scholarships for youth who are not parents could have been recorded as Education and Training (Line 6a2, 196), Prevention of Out-of-Wedlock Pregnancies (Line 6h, 196) or Other (Line 6m, 196). Still, with no way to disaggregate these expenditures this crosswalk represents the closest possible alignment between data from the two stable periods. A table outlining those transformations is included in Appendix A.

This conversion allowed us to append the data from the two stable periods—1997-2014 and 2015-2023—resulting in a single file containing data for the full period. We then appended the corresponding field name to the column names, which were up to this point the line numbers from ACF-196R (i.e. a column named “1” would be renamed “1. Awarded”). These names can also be seen in Appendix A.

### Tableau Variables

In its visualizations and reports, OFA references a consolidated set of metrics: Basic Assistance, Work, Education, & Training Activities, Child Care (Spent or Transferred), Program Management, Refundable Tax Credits, Child Welfare Services, Pre-Kindergarten/Head Start, Transferred to SSBG, Out-of-Wedlock Pregnancy Prevention, Non-Recurrent Short Term Benefits, Work Supports & Supportive Services, Services for Children & Youth, Authorized Solely Under Prior Law, Fatherhood & Two-Parent Family Programs, Other. Each of these metrics is equivalent to the sum of one or more lines in ACF-196R. These variables are present in Tableau-specific files that underlie the views in the expenditure data’s Tableau dashboard. The set of instructions used to create each variable can be seen in Appendix C.

We created two new variables for inclusion in the files used to generate Tableau dashboards: pct\_of\_tanf and pct\_of\_total. The first, pct\_of\_tanf, calculates the ratio of an expenditure category to the total funds available for TANF. The total funds available for TANF is calculated as the sum of Total Expenditures (24, 196R), Transfers to Child Care and Development Fund (CCDF) Discretionary (2, 196R), and Transfers to Social Services Block Grant (3, 196R). This sum is done within a given state, year, and funding level combination. The ratio is calculated by dividing the amount of expenditure in a category by the calculated total.

The second variable, pct\_of\_total, displays the percentage of the total expenditures in a category that can be attributed to the state or federal funding levels. For example, if in state A in year X the total expenditure in a category is $100 and the expenditure in that category at the federal and state funding levels is $55 and $45 respectively, then pct\_of\_total at the federal level will be 55% and pct\_of\_total at the state level will be 45%.

## Caseload Data

### Description of the Data

The Office of Family Assistance (OFA) collects data concerning Temporary Assistance for Needy Families (TANF) caseloads monthly. Annually, OFA compiles this data and publicly releases a dataset summarizing TANF caseloads. The TANF caseload data are currently published across three Microsoft Excel workbooks each fiscal year: a workbook containing TANF, or federal, caseload figures, a workbook containing Separate State Programs (SSP) and Maintenance of Effort (MOE), or state, caseload figures, and a workbook containing the total caseload (sum of TANF and SSP-MOE). These workbooks contain tables reporting monthly caseload figures, as well as tables reporting the average caseload for the fiscal year (October - September). The average number of families and individual recipients receiving TANF are reported separately. The longitudinal file we generate uses only the average caseloads for each fiscal year, in each file.

### Transformations

While the caseload data is already relatively consistent across time, some work is necessary to align the data for longitudinal use. We applied the following transformations to generate a longitudinal file:

* Merged family and recipient sheets – As noted in the description of the data, the average number of family and individual participants in TANF are reported separately. For our purposes, this is not necessary. Therefore we merge these two worksheets.
* Renamed columns – Column names are similar, but inconsistent across time. To ensure that columns always align, we assign them consistent names over time. Column names, definitions, and other metadata can be seen in Appendix XX.
* Standardize state names – We resolve some inconsistencies in State names throughout the files. For example:
  + Washington D.C. is always denoted “District of Columbia.”
  + The row totaling all states is always denoted “U.S. Total” to align with the expenditure data.
  + We correct a typo in “Montana” in at least one case.
  + Any characters indicating references to notes are removed (for example, “\*”, or numbers such as “1”).
  + We correct the mislabeling of Wisconsin in 2004.
* Remove notes and header rows.
* Round numeric values to the nearest 100th.

1. [State TANF Spending in FY 2015 | The Administration for Children and Families](https://www.acf.hhs.gov/ofa/data/state-tanf-spending-fy-2015) [↑](#footnote-ref-2)