

---

# Chapter 1: Producing Welfare Outcomes Data Files

A major goal of ASPE’s Welfare Outcomes grants is to ensure that the administrative and survey data used in grantees’ analyses are made available to qualified researchers and to ASPE, either through public use or researcher-restricted access data files. In considering how to meet this grant requirement, the Work Group concluded that grantees should produce data files using a standardized approach. Standardizing data files will help researchers understand, interpret, and analyze grantee data, and may promote data usage in ongoing welfare outcomes research. In creating a standardized approach to designing and producing these files, the Work Group also hopes to simplify this task for grantees.

With all these goals in mind, the Work Group has developed guidance to assist grantees in producing standardized data files. The following chapter provides grantees with an overview of this recommended approach.

## 1. File Structure

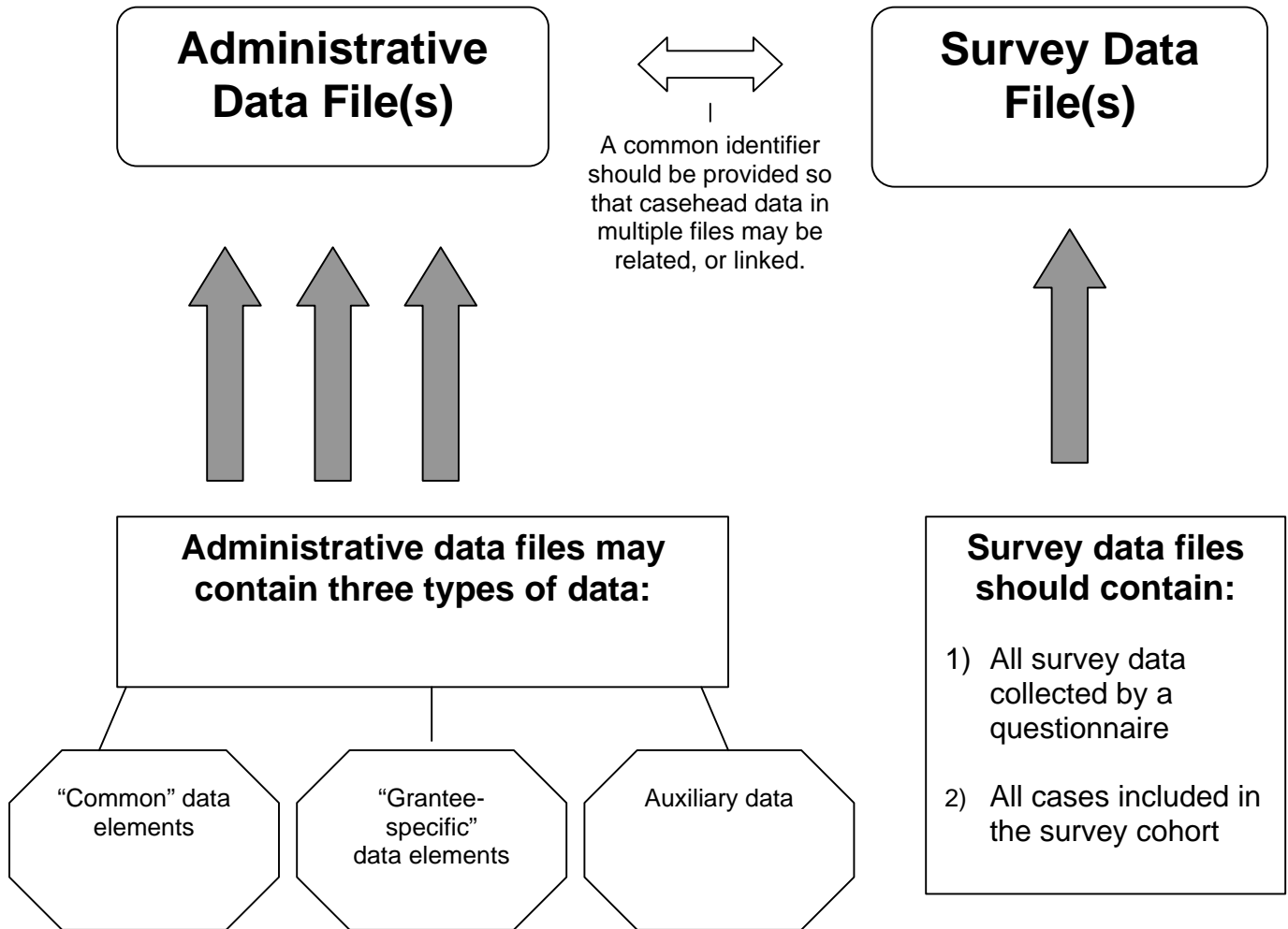
To give shape to the overall presentation of welfare outcomes data files, **the Work Group recommends that grantees follow a common file structure.** Organizing data files across grantees in a standardized way will help researchers gain a common understanding to any particular file’s content and context.

Grantees should consider producing survey and administrative data files as a set of separate files, across which individual cases may be related, or linked, using a common record identifier. As depicted in Figure 1.1: “Recommended Relational File Structure,” grantee **welfare outcomes data files may be reported as a set of two types of files: administrative data and survey data files.** The actual number of administrative and survey data files that grantees produce will be determined by the data characteristics, number of study cohorts, and the approach that presents the least burden.



An overview of the recommended file structure is illustrated in Figure 1.1, “Relational File Structure.”

**Figure 1.1: Recommended Relational File Structure**



---

## Administrative Data File Content

The Work Group recommends that all grantees produce at least one file that includes administrative data analyzed in their welfare outcomes research. At a minimum, this file should contain all data elements used in administrative data reports, including variables constructed for analysis and many of the raw data elements used in constructing these variables.

After evaluating administrative data used by grantees in their research and included in their reports, the Work Group identified three basic types of administrative data that may be included in grantee files:

- **Common data elements.**

The Work Group identified a **list of 18 common data items** that most grantees analyzed and reported. For these items, the Group developed a **uniform set of definitions and reporting guidelines** to help grantees report these data in a standardized manner (see Chapter 2). These guidelines provide standard definitions, variable names, and, where applicable, minimum time periods over which grantees should report monthly or quarterly administrative data. In some cases, grantees may not have included a particular common data element in their analysis; where possible, grantees should add this data element to their welfare outcomes data file.

- **Grantee-specific data elements.**

Grantee-specific data elements are those that were included in individual grantees' research but are not found across all grantees. Given the rich administrative data analyses conducted in welfare outcomes grants, many data elements are found in certain grantees' research, but not others, and are not included in the list of recommended common data elements. These grantee-specific data elements should also be included in welfare outcomes data files. **Examples of grantee-specific data elements** are provided in Chapter 2.



Lists and examples of common, grantee-specific, and auxiliary data elements are presented in Chapter 2, Figures 2.1, 2.2, and 2.3.

---

- **Auxiliary data elements.**

Auxiliary data elements consist of data on recipient characteristics and welfare outcomes that may not have been included in grantees' research, but would provide useful information to researchers.

Grantees are encouraged to provide as much administrative data as possible in their data files. These rich data sets will support ongoing in-depth welfare outcomes research. To assist grantees with identifying other useful administrative data that may be included in their files, the Work Group developed a **list of some auxiliary data elements that further describe recipient characteristics and other welfare outcomes** (see Chapter 2, Figure 2.3 for list). Grantees are encouraged to add these data to their files if the data are available, and if their addition will not pose a significant additional burden to creating the file.

### **Producing Multiple Administrative Data Files**



Instructions for producing multiple administrative data files are presented in Tip 1.1.

**Grantees may decide to produce more than one administrative data file** in cases where they find it easier or more efficient to do so, or when certain types of administrative data do not easily fit into a single rectangular file (see Tip 1.1).

If grantees choose to report administrative data in more than one data file, the Work Group recommends the following:

- File documentation should clearly describe the data included in each file and how these data were incorporated into grantee analysis and reporting.
- Unique record identifiers should be assigned to particular cases in such a way that data can be linked across administrative and survey data files.
- If data for multiple study cohorts are reported in separate files, the same common, grantee-specific, and auxiliary data elements should be reported for each cohort, to the extent possible.

---

## ✓ Tip 1.1: Creating Multiple Administrative Data Files

Grantees may choose to produce more than one welfare outcomes administrative file when existing data do not easily fit within the proposed structure, OR when producing multiple files will decrease grantee burden.

For example, grantees may prefer to produce multiple administrative data files under the following scenarios:

- ❑ Analysis includes individual-level data, or records, within a case unit that have not been appended to casehead records to form one record per case unit (e.g., Medicaid program data where information for each child on the case may be reported as a separate record).
- ❑ Research involved multiple study cohorts from different time periods or with different study groups. For example, multiple study cohorts from different time periods may result in a linked file with a large number of unpopulated data fields for monthly or quarterly data. Also, if analysis included different study groups such as leavers and long-term recipients, it may be easier to report administrative data items for these groups in separate files.
- ❑ More than one file was constructed for analysis and reporting.

If grantees choose to produce multiple administrative data files, they should:

- ❑ Clearly document the types of data included in each file and the rationale behind how the files were constructed.
- ❑ Assign record identifiers to particular cases in such a way that data may be linked across survey and administrative data files.
- ❑ Report the same common, grantee-specific, and auxiliary data elements for separate study cohorts contained in multiple data files, to the extent possible.



Elements that go into a survey data file are itemized in Chapter 2, Tip 2.3.

## Survey Data Files

Grantees should make a **survey data file** available for secondary analysis. Where grantees' research includes multiple surveys, a separate survey data file should be produced for each survey that has been conducted. Survey data files should include:

- All survey data collected by a questionnaire. Limited exceptions to this rule are listed in Chapter 2.
- A separate record for all individuals who were included in the survey sample. This includes both survey respondents and non-respondents.
- Variables that were constructed for analysis from survey raw data.
- Survey administration items necessary for researchers to interpret and use the survey data (e.g., survey mode, final survey dispositions).

**A more detailed description of the items to be included in survey data files** is provided in Chapter 2.

## 2. File Format

**Administrative and survey data files available to researchers for secondary analysis should be produced using an ASCII fixed-field format** which is easily read and used by most statistical and analytical software packages. A record layout should be included in data file documentation.



See Chapters 3 and 4 for further information on documenting study cohorts and selection events.

---

### 3. Identification of Selection Month or Quarter

**Clear identification of the study cohort and the selection period is critical to future researchers' understanding of grantee welfare outcomes data files.**

In general, all grantee studies track welfare outcomes as occurring “x” months or quarters before or after a particular selection event (e.g., TANF exit or application). This selection event defines or “triggers” entry into a particular study cohort. For example, study cohorts in TANF leavers studies are defined by the month or quarter in which the recipient exits the TANF program. In studies with individuals who apply for or are diverted from receiving TANF benefits, an analogous event, such as the month of application, generally results in selection for a particular study cohort.

Given the large number of ASPE-funded leavers studies, this guide generally refers to the selection time period as the “exit” month. However, for those grantees studying different population groups (e.g., TANF applicants or divertees), the term “exit” month should be understood to mean the selection month used in their particular study.

For example, grantees conducting welfare outcomes studies with TANF leavers have chosen to **define exit month one of two ways:**

- The month in which an individual left TANF (i.e., the last month of benefit receipt).
- The first full month after an individual left TANF (i.e., the first month after case closure when no benefits were received).

Grantee file documentation for leavers studies should **clearly note which exit month definition was selected and the exit period used** (e.g., the exit months or quarters analyzed). File documentation for diversion and other studies also should clearly note the base month or quarter selected and the time periods analyzed in the study.

Some grantees have selected multiple study cohorts with different exit months. Administrative and survey data for these cohorts may be in one or more files. Where data are included in one file, grantees should include a flag variable that notes which cases were included in specific study cohorts. Alternatively, if study cohorts are reported as separate data files, file documentation should clearly note where researchers could locate data for a specific cohort.

---

## 4. Unit of Analysis

Grantee welfare outcomes studies with TANF leavers have identified the casehead as the primary unit of analysis. In most administrative and survey data files, file records, or rows, should correspond with a particular casehead. However, in certain types of administrative data, **the unit of analysis, or file record, may be something other than the casehead.** For example, certain data files may list each child or family member as a separate record. In these situations, grantees may report these data one of two ways:

- As a separate file with a defined variable that allows the data to be linked, or related, to casehead data contained in other files.
- By appending these data to casehead records, creating a new set of variables associated with each child or family member case.

Grantees' administrative data file documentation should **indicate where the units of analysis and file records do not correspond with a casehead,** and should specify the file's unit of analysis.



---

## 5. Creating Record Identifiers



Tip 1.2, below, shows further details on creating record identifiers.

**All cases included in survey and administrative data files should be assigned a unique record identifier.** The Work Group recommends that these identifiers be created using a combination of state FIPS codes and a randomly assigned record number (see Tip 1.2).

If a particular case is located in more than one administrative or survey data file, the same identifier should be assigned to this case in all files. Keeping identifiers consistent will allow case-level data to be related, or linked, across files. Even if grantees have not chosen similar study cohorts for their administrative and survey data analysis, they should still assign unique identifiers for individual cases and apply them consistently in both administrative and survey data files where a particular case coincidentally appears in both sources. File documentation should note where researchers should not expect a one-to-one match between cases included in administrative and survey data files.

### ✓ Tip 1.2: Creating Record Identifiers

Grantees may construct a 10-digit record identifier using a combination of state FIPS codes and a random number for survey and administrative data file records.

- ❑ State Identifier (Columns 1-2): State FIPS code.
- ❑ Random Case Identifier (Columns 3-10): A random number identifier that is not correlated with any individual characteristic. For example, files should not be sorted by a specific field and then sequentially numbered.

Individual cases listed in more than one data file should be assigned the same record identifier.



The Grantee Toolkit contains a sample SAS program that may be used to assign random case identifiers.

---

## 6. Creating Variable Names

The Work Group recommends following a standard approach for creating **variable names**. Consistent variable naming conventions, especially in the case of administrative data, will help researchers to understand and use welfare outcomes data.

### Administrative Data



Variable names for common, grantee-specific and auxiliary administrative data items are presented in Chapter 2, Figures 2.1, 2.2, and 2.3.

The Work Group developed a list of variable names that grantees can use for the common, and some of the “grantee-specific” or auxiliary administrative data elements (see Chapter 2). For all other administrative data elements included in the grantee welfare outcomes files, the Work Group recommends the following approach.

Administrative data are generally reported either as a single occurrence (e.g., gender, and number of adults on case at time of “exit”) or for multiple time periods (e.g., monthly/quarterly and before and after “exit”). **Administrative data reported as a single occurrence should be assigned a descriptive name that is eight characters in length.**

If a grantee has tracked a particular administrative data element over time (e.g., quarterly earnings), the data file should include a separate variable for each interval. **Variable names for administrative data elements that recur over time should have the following three components:**

- A variable name.
- A reference to whether the data element reflects the time period before or after TANF “exit,” or other selection event (e.g., TANF application).
- An indicator as to whether these data recur monthly or quarterly.
- A reference to the number of months or quarters prior to or following the TANF “exit,” or other selection event (e.g., TANF application).



An overview of naming recurring administrative variables is presented in Figure 1.2.

Instructions for constructing monthly and quarterly administrative data variable names are given in Tips 1.3 and 1.4.

Figure 1.2: “Recommended Approach to Naming Recurring Administrative Data Variables” illustrates how variable names may be constructed. Examples of how this approach may be applied are provided in Tips 1.3 and 1.4.

**Figure 1.2: Recommended Approach to Naming Recurring Administrative Data Variables**

<b>Columns 1-4: Variable Name</b>	<b>Column 5: Reference to Selection Event</b>	<b>Column 6: Quarterly or Monthly Data</b>	<b>Columns 7-8: Time Period Reference</b>
<p>Variable names for “common” and “auxiliary” administrative data elements are provided in Chapter 2.</p> <p>Grantees should create a descriptive four-character variable name for each grantee-specific and auxiliary data item.</p>	<p>B = Prior To TANF “exit” or other selection event</p> <p>A = Post TANF “exit” or other selection event</p> <p>TANF “exit” or other selection event should be clearly defined in the administrative data file documentation.</p>	<p>M = Monthly data</p> <p>Q = Quarterly data</p>	<p>Specific month or quarter prior to or following TANF “exit” or other selection event: 01, 02, 03, 04, etc.</p> <p>For example, data from the month or quarter of TANF “exit” should be noted as AM00 or AQ00.</p>

### Survey Data

**Survey data variable names should correspond with their labels in the survey questionnaire.** Constructed survey variables may be given a variable name that is **eight characters in length**. Survey data file documentation should list constructed survey variables.

---

## ✓ Tip 1.3: Constructing Variable Names for Monthly Administrative Data

For the purpose of this example, assume we are modeling food stamp receipt during the month prior to the TANF exit month.

**Step 1: Assign variable name**

The first four columns should include a descriptive variable name.

*For our example case, the description could be "FOOD."*

**Step 2: Assign selection event reference**

The fifth column should include a reference to whether the data are from before or after TANF exit or other selection event: "B" refers to the period before TANF exit, and "A" refers to the time period following TANF exit.

*For our example case, the data element comes from the month prior to the TANF exit month. Accordingly, the next column would be "B."*

**Step 3: Identify the data type**

The sixth column refers to whether the data are monthly or quarterly: "M" refers to monthly, and "Q" refers to quarterly.

*For our example case, the data are monthly and would be noted as "M."*

**Step 4: Include time period reference**

The seventh and eighth columns are a numeric value that refers to the number of months or quarters before or after TANF exit or other selection event. Numeric values should always be reported in two digits (e.g., 01, 02, 03, etc.).

*For our example case, the time period reference would be 01 for the month prior to the TANF exit month.*

**Put it all together**

*Taken all together, the variable name for this data element is: FOODBM01*

---

## ✓ Tip 1.4: Constructing Variable Names for Quarterly Administrative Data

For the purposes of this example, assume we are modeling unemployment earnings during the quarter following the TANF exit month.

- |  |   |
|--|---|
| <input type="checkbox"/> <b>Step 1: Assign variable name</b>             | The first four columns should include a descriptive variable name.<br><br><i>For our example case, the description could be "EARN."</i>   |
| <input type="checkbox"/> <b>Step 2: Assign selection event reference</b> | The fifth column should include a reference to whether the data are from before or after TANF exit or other selection event: "B" refers to the period before TANF exit, and "A" refers to the time period following TANF exit.<br><br><i>For our example case, the data element is from the first month following the TANF exit month. Accordingly, the next column would be "A."</i> |
| <input type="checkbox"/> <b>Step 3: Identify the data type</b>           | The sixth column refers to whether the data are monthly or quarterly: "M" refers to monthly, and "Q" refers to quarterly.<br><br><i>For our example case, the data are quarterly and would be noted as "Q."</i>   |
| <input type="checkbox"/> <b>Step 4: Include time period reference</b>    | The seventh and eighth columns are a numeric value that refers to the number of months or quarters before or after TANF exit or other selection event. Numeric values should always be reported in two digits (e.g., 01, 02, 03, etc.).<br><br><i>For our example case, the time period reference would be 01 for the quarter after the TANF exit month.</i>                          |
| <input type="checkbox"/> <b>Put it all together</b>                      | <i>Taken all together, the variable name for this data element is: EARNAQ01</i>   |