

Human Services Domain Data Harmonization Procedures

Human Services Domain
Document Version 2.4
November 29, 2016

Administration for Children and Families
Office of Child Support Enforcement
330 C St. SW
Washington, DC 20024

TABLE OF CONTENTS

Executive Summary	1
1. PRELIMINARY DATA HARMONIZATION TASKS	2
1.1 Task: Select the Data Elements to Harmonize.....	2
1.2 Task: Obtain Documentation for the Selected Elements	2
1.3 Task: Analyze Reference Source Documents.....	2
1.4 Task: Data Harmonization Spreadsheet.....	2
2. DATA ELEMENT ATTRIBUTE DESCRIPTIONS	3
2.1 ID	3
2.2 Data Element Name	4
2.3 Human Services Domain Standard Name.....	5
2.4 Label	5
2.5 Definition	6
2.6 Authoritative Source	6
2.7 Length [L] and Type [T].....	7
2.8 Business Rules	8
2.9 Value Domain	8
2.10 Issues/Comments	9
2.11 Mapping	9
2.12 Validated By/Validation Date.....	10
3. EXTRACTION OF DATA FOR HARMONIZATION	11
3.1 Extract and Collect Data Elements	11
3.1.1 Extracting Metadata from Forms and Form Instructions.....	11
3.1.2 Extracting Data from Electronic Data Exchange.....	14
3.2 Initial Validation	16
3.3 Identify Anomalies.....	17
3.4 Integrate and Collate Data Elements.....	18
3.5 Final Validation	19
Appendix A: Examples	20

Executive Summary

The Human Services Domain Data Harmonization Procedures document details the process to better understand data, identify opportunities to reduce duplicated data, and develop common data standards within the Human Services Domain. Standardizing data elements reduces ambiguity, provides consistency, and reduces the potential for errors. Data harmonization is the first step in the process of developing data standards.

Data elements are defined by two primary attributes: a description and a value. For example, the data element Social Security Number would have a:

- Description: a unique number assigned to a person by the Social Security Administration, and a
- Value: a value would be 080-34-5678.

The example of Social Security Number is also defined by its attribute name and by other descriptive attributes. These attributes are referred to as metadata, data that describes data.

This document also examines current data element definitions, data element related Business Rules and descriptions, and applies a process to provide consistent data element definitions and descriptions for Human Services Domain data. Data harmonization refines the data elements into a standardized form through the following steps:

1. Extract the data attributes
2. Perform an initial validation
3. Identify and resolve anomalies
4. Integrate and collate the extracted data elements
5. Perform a final validation, and migrate the spreadsheet to the Data Standards Registry

Data harmonization can be applied to data elements used in data exchanges, databases, and to data fields within forms, i.e. electronic and paper respectively. Harmonizing data within forms requires additional steps, which are described in separate sections.

- Section 1 provides an overview of the steps in the data harmonization process.
- Section 2 details the initial data harmonization process for selecting data elements to be included and describes each of the attributes to be extracted into a spreadsheet template.
- Section 3 describes the step-by-step process for preparing the data harmonization spreadsheet, in preparation for the migration of the spreadsheet contents into a Data Standards Registry (Metadata Repository).

A critical part of the data harmonization process is providing a standardized name for the data elements. For supplementary guidelines, please refer to *Appendix A: Examples* and the *Human Services Domain Data Standards Naming Conventions and Controlled Vocabulary*.

1. PRELIMINARY DATA HARMONIZATION TASKS

The following preliminary tasks should be done prior to starting the data harmonization process.

1.1 Task: Select the Data Elements to Harmonize

The first step in the process is to identify the data exchanges, databases, or forms from which to extract data elements.

1.2 Task: Obtain Documentation for the Selected Elements

Collect all source reference documents. References may include:

- Program forms
- Guidance documents
- Data design documents
- Data definition documents
- Forms and their instructions
- Glossaries
- Data dictionaries

Various sources may contain different or additional information about the elements to be harmonized.

1.3 Task: Analyze Reference Source Documents

Review the documents, identifying key attributes of the fields or elements as discussed in section 3.1. An understanding of the data harmonization spreadsheet prior to analyzing the reference material is helpful in identifying the attributes.

1.4 Task: Data Harmonization Spreadsheet

Using a data harmonization spreadsheet template discussed in section 2, enter the name for the program or form to be harmonized.

2. DATA ELEMENT ATTRIBUTE DESCRIPTIONS

The data elements attributes, semantic information or metadata, are extracted from the reference documents. It is then entered into the data harmonization spreadsheet template, which is and must be a Microsoft Excel spreadsheet as shown below.

Source Name													Date													Extracted by													Version number												
ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/ Comments	Mapping	Validated By/ Validation Date																																							

In the top row of the spreadsheet enter the **Source Name** [the program (data exchange) or form name from which you are extracting the data], the **Date**, **Extracted by** [your name], and a **Version Number**. The first version should be 0.1.

A description of each column heading (metadata) is provided in the following sections. The convention that is followed to describe the column headings is as follows:

Title – Follows the ‘*Section *.**’

Use – Explains the usage

Note – Provides important information to the reader regarding the attribute – *Optional*

Requirement (or Restriction) – States the requirements or any restrictions for this attribute field – *Optional*

Example – Provides an example. *Appendix A* contains additional examples where clarification may be necessary.

2.1 ID

Use: The ID provides an unambiguous identifier in order to differentiate data elements.

Note: The unique ID should be assigned after all metadata has been extracted from the source documents.

Requirement: All data elements in the data harmonization spreadsheet must have a unique identifier. The ID must contain a reference to the source program name, the input/output file or form name from which it was extracted, and a unique sequence number.

Example: As shown below, the data element named Benefit Amount has an unique identifier. The unique ID, SWA-UI-OD-19, assigned is broken down into the following:

- SWA - State Workforce Agency
- UI - Unemployment Insurance

- **OD** - Output Detail Record
- **19** – the 19th data element extracted

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/ Comments	Mapping	Validated By/ Validation Date
SWA-UI-OD-19	Benefit Amount	Unemployment Insurance Benefit Amount	Benefit Amount	The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.	Statue	11	N	Values are 0000000000 through 9999999999 without decimal. This field is whole dollars only.	Values are 0000000000 through 9999999999 without decimal.	ANOMALY: Some dollar values are expressed with an implied decimal point.	SWA UI Output Detail Record FCR UIFSA General Testimony	Validated by: John Doe 5/12/2006

2.2 Data Element Name

Use: Provides a unique and contextually meaningful name for the data element.

Note: The Data Element Name is extracted exactly as it appears in the source document. Later, during the validation process, additional descriptors may be added. An example appears in *Appendix A*.

Requirement: The Data Element Name is the name extracted from the source document, if specified. A name can be extracted from an input/output record specification, data dictionary, or a form’s instructions. Each name must be unique within a program or form. Once the data element name has been extracted from the form, it is important to examine the purpose or meaning of the name in the context in which it appears on the form.

Example: As shown below, the data element name **Benefit Amount** was extracted from the **SWA**.

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/ Comments	Mapping	Validated By/ Validation Date
SWA-UI-OD-19	Benefit Amount	Unemployment Insurance Benefit Amount	Benefit Amount	The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.	Statue	11	N	Values are 0000000000 through 9999999999 without decimal. This field is whole dollars only.	Values are 0000000000 through 9999999999 without decimal.	ANOMALY: Some dollar values are expressed with an implied decimal point.	SWA UI Output Detail Record FCR UIFSA General Testimony	Validated by: John Doe 5/12/2006

2.3 Human Services Domain Standard Name

Use: Provides a Human Services Domain name to a data element.

Note: The *Human Services Domain Data Standards Naming Conventions and Controlled Vocabulary* document details the Human Services Domain standard naming conventions. The document covers such issues as composition, sequence, controlled vocabulary, and conventions. It will assist you in generating a valid Human Services Domain Standard Name.

Requirement: Data elements must be assigned a Human Services Domain name. Reference the *Human Services Domain Data Standards Naming Conventions and Controlled Vocabulary* for further guidance.

Example:

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/ Comments	Mapping	Validated By/ Validation Date
SWA-UI-OD-19	Benefit Amount	Unemployment Insurance Benefit Amount	Benefit Amount	The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.	Statue	11	N	Values are 0000000000 through 9999999999 without decimal. This field is whole dollars only.	Values are 0000000000 through 9999999999 without decimal.	ANOMALY: Some dollar values are expressed with an implied decimal point.	SWA UI Output Detail Record FCR UIFSA General Testimony	Validated by: John Doe 5/12/2006

2.4 Label

Use: Identifies and names data field.

Note: Titles that appear on a form page number, section headers (e.g., ‘IV: Additional Information’) within the form, and the OMB reference number are extracted as Labels and must be included in the spreadsheet. These labels typically do not require other metadata attributes such as a definition. Labels must also be repeated as many times as they appear on the form (as in the case of a form with several pages).

Requirement: If the source document is a form then a label or question typically appears for each data field. The label or question should be extracted and placed in the Label column of the spreadsheet exactly as it appears on the form. Example: Benefit Amount

Human Services Domain Data Harmonization Procedures

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/ Comments	Mapping	Validated By/ Validation Date
SWA-UI-OD-19	Benefit Amount	Unemployment Insurance Benefit Amount	Benefit Amount	The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.	Statue	11	N	Values are 0000000000 through 9999999999 without decimal. This field is whole dollars only.	Values are 0000000000 through 9999999999 without decimal.	ANOMALY: Some dollar values are expressed with an implied decimal point.	SWA UI Output Detail Record FCR UIFSA General Testimony	Validated by: John Doe 5/12/2006

2.5 Definition

Use: Contains the data elements or data field definition extracted from the source document(s).

Note: Do not include rules or format information.

Requirement: Every definition must reference an Authoritative Source. For more details on the definition rules, please refer to the *Human Data Element Naming Conventions and Controlled Vocabulary*.

Example: The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/ Comments	Mapping	Validated By/ Validation Date
SWA-UI-OD-19	Benefit Amount	Unemployment Insurance Benefit Amount	Benefit Amount	The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.	Statue	11	N	Values are 0000000000 through 9999999999 without decimal. This field is whole dollars only.	Values are 0000000000 through 9999999999 without decimal.	ANOMALY: Some dollar values are expressed with an implied decimal point.	SWA UI Output Detail Record FCR UIFSA General Testimony	Validated by: John Doe 5/12/2006

2.6 Authoritative Source

Uses: Identifies the reference source for the data element definition.

Note: Authoritative sources include statutory and regulatory citations, and citations from Human Services Domain documents. Authoritative sources can also be Human Services Domain subject matter experts or

long-standing child support convention, as well as external references, such as FIPS. The precedence of these authoritative sources, in descending order, is:

1. Statute
2. Regulation
3. Black’s Law Dictionary
4. Action Transmittal/ Policy Interpretation
5. Dear Colleague Letter (DCL)
6. Subject Matter Expert
7. Convention

Statutes are the preferred source of data element name definitions. However, because of their specialized usage and need, many definitions will not have references in a statute or regulation. Research other sources for the appropriate definitions.

Requirement: Every definition must reference an Authoritative Source.

Example: Merriam Webster’s Dictionary for a Convention.

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/ Comments	Mapping	Validated By/ Validation Date
SWA-UI-OD-19	Benefit Amount	Unemployment Insurance Benefit Amount	Benefit Amount	The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.	Statue	11	N	Values are 00000000000 through 99999999999 without decimal. This field is whole dollars only.	Values are 00000000000 through 99999999999 without decimal.	ANOMALY: Some dollar values are expressed with an implied decimal point.	SWA UI Output Detail Record FCR UIFSA General Testimony	Validated by: John Doe 5/12/2006

2.7 Length [L] and Type [T]

Use: Describe the length in bytes of the data element name and the type (alphabetic, numeric, or alphanumeric).

Note: On most paper forms length and type are not defined. These entries should be left blank; the Length and Type columns are populated during the final validation step.

Requirement: N/A

Example: In the example below, Benefit Amount has a length of 11 bytes and a numeric type. These values were extracted from the SWA record layout document.

Human Services Domain Data Harmonization Procedures

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/ Comments	Mapping	Validated By/ Validation Date
SWA-UI-OD-19	Benefit Amount	Unemployment Insurance Benefit Amount	Benefit Amount	The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.	Statue	11	N	Values are 00000000000 through 99999999999 without decimal. This field is whole dollars only.	Values are 00000000000 through 99999999999 without decimal.	ANOMALY: Some dollar values are expressed with an implied decimal point.	SWA UI Output Detail Record FCR UIFSA General Testimony	Validated by: John Doe 5/12/2006

2.8 Business Rules

Use: Defines possible values and formatting for a data element.

Note: Sometimes the Business Rules are provided in the source documentation as definitions. For more information on this topic and examples see section 3.3.

Restriction: Business rules may vary depending on the program, data exchange, or context in which they are used.

Example: This field is whole dollars only.

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/ Comments	Mapping	Validated By/ Validation Date
SWA-UI-OD-19	Benefit Amount	Unemployment Insurance Benefit Amount	Benefit Amount	The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.	Statue	11	N	Values are 00000000000 through 99999999999 without decimal. This field is whole dollars only.	Values are 00000000000 through 99999999999 without decimal.	ANOMALY: Some dollar values are expressed with an implied decimal point.	SWA UI Output Detail Record FCR UIFSA General Testimony	Validated by: John Doe 5/12/2006

2.9 Value Domain

Use: Defines the valid values of a data element.

Note: N/A

Restriction: Valid values should not include business rules.

Human Services Domain Data Harmonization Procedures

Example: Values are 0000000000 through 9999999999 without decimal.

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/Comments	Mapping	Validated By/Validation Date
SWA-UI-OD-19	Benefit Amount	Unemployment Insurance Benefit Amount	Benefit Amount	The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.	Statue	11	N	Values are 00000000000 through 99999999999 without decimal. This field is whole dollars only.	Values are 0000000000 through 9999999999 without decimal.	ANOMALY: Some dollar values are expressed with an implied decimal point.	SWA UI Output Detail Record FCR UIFSA General Testimony	Validated by: John Doe 5/12/2006

2.10 Issues/Comments

Use: Documents any observations, assumptions, suggestions or questions that arise during data extraction.

Note: Issues and comments may be added to the data harmonization spreadsheet at any time during the extraction or validation steps. Issues and comments will be stored in the Data Standards Registry when the harmonized data is transferred. Therefore, it is important that this data be retained even if an issue or question is resolved or answered.

Restriction: Documentation of all rationales, commentary, or suggestions pertaining to a particular data element.

Example: ANOMALY: Some dollar values are expressed with an implied decimal point.

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/Comments	Mapping	Validated By/Validation Date
SWA-UI-OD-19	Benefit Amount	Unemployment Insurance Benefit Amount	Benefit Amount	The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.	Statue	11	N	Values are 00000000000 through 99999999999 without decimal. This field is whole dollars only.	Values are 00000000000 through 99999999999 without decimal.	ANOMALY: Some dollar values are expressed with an implied decimal point.	SWA UI Output Detail Record FCR UIFSA General Testimony	Validated by: John Doe 5/12/2006

2.11 Mapping

Use: Lists all the programs or forms that contain a specific data element.

Note: N/A

Requirement: N/A

Example: The data element appears in the following programs/forms:

- SWA UI Output Detail Record
- FCR
- UIFSA General Testimony, OMB 0970-0085, exp. 01/31/2008.

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/ Comments	Mapping	Validated By/ Validation Date
SWA-UI-OD-19	Benefit Amount	Unemployment Insurance Benefit Amount	Benefit Amount	The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.	Statue	11	N	Values are 0000000000 through 9999999999 without decimal. This field is whole dollars only.	Values are 0000000000 through 9999999999 without decimal.	ANOMALY: Some dollar values are expressed with an implied decimal point.	SWA UI Output Detail Record FCR UIFSA General Testimony	Validated by: John Doe 5/12/2006

2.12 Validated By/Validation Date

Use: Validation phase.

Note: This column will be filled out by the validator.

Requirement: N/A

Example: Validated by: John Doe 5/12/2006.

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Business Rules	Value Domain	Issues/ Comments	Mapping	Validated By/ Validation Date
SWA-UI-OD-19	Benefit Amount	Unemployment Insurance Benefit Amount	Benefit Amount	The gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.	Statue	11	N	Values are 0000000000 through 9999999999 without decimal. This field is whole dollars only.	Values are 0000000000 through 9999999999 without decimal.	ANOMALY: Some dollar values are expressed with an implied decimal point.	SWA UI Output Detail Record FCR UIFSA General Testimony	Validated by: John Doe 5/12/2006

3. EXTRACTION OF DATA FOR HARMONIZATION

This section describes the process for extracting data elements and data fields from source documents.

There are six steps in the extraction process. Specific instructions for extracting data fields from forms are included in each step. The steps are:

1. Extract and Collect Data Elements and Data Fields
2. Perform an Initial Validation
3. Identify and Resolve Anomalies
4. Integrate and Collate the Data Elements
5. Perform a Final Validation
6. Migrate Data Elements to the Data Standards Registry

3.1 Extract and Collect Data Elements

The first step in the harmonization process is to extract the metadata and enter it into the Data Harmonization Spreadsheet template.

During the extraction step, the metadata must be extracted from the source documentation exactly as it appears. This includes spelling or grammatical errors. Extract the Data Element Name, Definition [if provided], and Business Rules [if provided] without adding or editing any text.

3.1.1 Extracting Metadata from Forms and Form Instructions

Data fields on forms appear as unstructured data. In the extraction process some structure must be given to the data without changing the meaning or semantics of the data. In addition, the data fields on the form are typically grouped together and are presented in a logical sequence to help the person filling out the form. Some of this detail is lost through the extraction process. In the data harmonization process the goal is to extract data elements as stand-alone entities. There is currently no facility to capture the business rules governing the data exchanges or instructions for forms. However, some of the semantics for these data elements are contained in the instructions and implied in the layout of the form.

Extracting data fields from forms requires additional preplanning. Forms are typically accompanied by instructions for filling them out. The instructions may contain definitions, business rules, and the relationships between data elements.

For example the UIFSA Data Sheet contains a data field called **Social Security Number(s)** [Data Element Name] and the instructions state, “Provide ‘Social Security Number(s)’, if known; this information is vital”. The ‘(s)’ implies an unlimited (unbounded) list of Social Security Numbers.

General rules for extracting metadata from forms:

- When data fields are repeated on a form, such as “name”, it is necessary to differentiate each instance. A prefix word can be assigned to each data element’s name. Example: **“Parent” First Name**, **“Child” First Name**.
- The title of a form is extracted as a label. Example: **Child Support Enforcement Transmittal #1 – Initial Request**.

- If the form is divided into sections, each section name is extracted as a label. Example: **VII Additional Case Information.**
- Single Data Field Extracted as Two or More Data Elements: An example is “Height”. In the instructions for UIFSA Locate Data Sheet, height is expressed in feet and inches. Therefore two extracted data elements are required: **Height Feet** and **Height Inches.**
- Extracting Business Rules from Form Instructions: Identify potential rules from the instructions that apply to the data element. Only capture business rules that are directly relevant to a specific data element.
- Extracting an Address Data Field: When extracting an address data field from a form, examine how it is represented on the form or in the instructions. If the form does not explicitly ask for each component for an address, use the following format:

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Value Domain	Business Rules	Issues/Comments	Mapping	Validated By/Validation Date
			<u>Address Line 1</u> <u>City</u> <u>State</u> <u>ZIP Code</u>									

For example: **Last Known Address** in the UIFSA Locate Data Sheet. This data field is not specified in more detail on the form or in the instructions. The data element should be extracted as follows:

- Last Known Address Line 1**
- Last Known Address City**
- Last Known Address State**
- Last Known Address ZIP Code**

However, if the form does specify the details of the address, extract as a representation of the address. For example, the data field on the form **Employer (Street, City, State, and Zip)** would be extracted as follows:

- Employer Street**
- Employer City**
- Employer State**
- Employer Zip**

- Extracting Person Name: A person (mother, father, caretaker, child, petitioner, etc.) on a form should always be extracted and extracted as follows:

- Full Name**
- First Name**
- Middle Name**

Last Name

For example: Petitioner should be extracted as follows:

- Petitioner Full Name**
- Petitioner First Name**
- Petitioner Middle Name**
- Petitioner Last Name**

- Multiple Instances of the Same Data Field (Unbounded): A data field such as “Dependent Child(ren)” on a form implies the data element can appear multiple times for as many children as there are on the case. The extracted data element should appear as follows:

Data Element Name: **Dependent Child Full Name (N)**. The (N) signifying an unlimited (unbounded) number.

- Multiple Instances of the Same Name (Bounded): Sometimes, the form or the instructions indicate a specific number of times a data element can appear. For example, if a data field appears 3 times, the data field should be extracted as follows:

- Petitioner Full Name (1)
- Petitioner Full Name (2)
- Petitioner Full Name (3)

- Be as expansive as possible. Example: “Race”. A separate data element should be extracted for each instance. For example:

- Race Black (non-Hispanic)**
- Race Hispanic**

- Checkboxes: If a data field on the form is a checkbox, put the following note in the spreadsheet Business Rules column: “Checkbox”.

- Labels: Labels should be extracted as individual data elements. Form titles are labels, as are all individual headings, names, page numbers, OMB number, etc.

- Example: Locate Data Sheet. A case identified can be either a IV-D or a non-IV-D case. The IV-D case is broken down into five categories, one of which is TANF. ‘IV-D Case’ and ‘TANF’ would be extracted as follows:

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Value Domain	Business Rules	Issues/ Comments	Mapping	Validated By/ Validation Date
			IV-D Case									
			IV-D Case TANF						Checkbox			

- **Business Rules/Definitions:** It is sometimes difficult to differentiate between a Definition and a Business Rule, as expressed in the forms instructions. The important thing to remember is that it be extracted to the spreadsheet – it can be reviewed and moved if necessary during the Validation stage. See *Human Services Domain Data Standards Naming Conventions and Controlled Vocabulary*.

3.1.2 Extracting Data from Electronic Data Exchange

Source documents used for data harmonization come in many form. Below are examples of sources:

- Data dictionaries
- Record layouts
- Documentation

This section illustrates how these source documents typically appear and provides tips for extracting metadata from them.

Data Dictionaries

The *FCR Interface Guidance Document* provides a data dictionary which is a source of metadata. The *Interface Guidance Document* has the following entry for the **State of Hire** data element:

Name: STATE OF HIRE
Type: Output Field
Condition: Conditional for the following output record: • FCR NDNH
Locate/Proactive Match Response Record
Length: 2
Format: Alphanumeric
Values: The alphabetic FIPS state or territory code or spaces
Description: The state or territory where the employee was hired.

When extracted for the data harmonization spreadsheet, **State of Hire** appears as follows:

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Value Domain	Business Rules	Issues/ Comments	Mapping	Validated By/ Validation Date
XX	State of Hire			The state or territory where the employee was hired.		2			The alphabetic FIPS state or territory code or spaces		FCR	

Record Layouts

The **Quarterly Wage Employee Wage Amount** data element appears in the record layout from the proposed IRS QW Output Detail Record:

CHART 5: PROPOSED IRS EITC QW OUTPUT DETAIL RECORD				
Field Name	Location	Length	Alpha/Numeric	Comments
QW Employee Wage Amount	408-418	11	A/N	This field will contain the information as provided from the QW record submitted to the NDNH.

The extracted metadata is shown below.

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Value Domain	Business Rules	Issues/Comments	Mapping	Validated By/Validation Date
YY	QW Employee Wage Amount					11			This field will contain the information as provided from the QW record submitted to the NDNH.	Proposed IRS EITC QW Output Detail Record	Data Access IRS Record	

Forms Instructions

The following process steps apply to the extraction of data fields from forms.

The data field **IV-D Case Medicaid Only** appears on the UIFSA Locate Data Sheet form. A Business Rule is extracted from the instructions for the Locate Data Sheet: “The obligee’s family receives Medicaid, but does not receive TANF (IV-A) cash payments”. This Business Rule is added [verbatim] from the instructions along with a unique ID, Mapping, and a note on implied optionality in the Issues/Comments column.

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Value Domain	Business Rules	Issues/Comments	Mapping	Validated By/Validation Date
ZZ	IV-D Case Medicaid Only								“The obligee’s family receives Medicaid, but does not receive TANF (IV-A) cash payments”.	Implies: IF Medicaid, THEN no TANF	UIFSA Locate Date Sheet OBM No. 0970-0085	

Once data from the various source documents is extracted, combine the rows in the spreadsheet. **State of Hire, QW Employee Wage Amount, and IV-D Case Medicaid Only**, would appear in the spreadsheet as:

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Value Domain	Business Rules	Issues/ Comments	Mapping	Validated By/ Validation Date
XX	State of Hire			The state or territory where the employee was hired		2	A/N		The alphabetic FIPS state or territory code or spaces		FCR	
YY	QW Employee Wage Amount					11	A/N		This field will contain the information as provided from the QW record submitted to the NDNH.	Proposed IRS EITC QW Output Detail Record	Data Access IRS Record	
ZZ	IV-D Case Medicaid Only								"The obligee's family receives Medicaid, but does not receive TANF (IV-A) cash payments."	Implies: IF Medicaid, THEN no TANF	UIFSA Locate Date Sheet OBM No. 0970-0085	

3.2 Initial Validation

The Initial Validation stage verifies that the Data Element Name, the Definition, and the Business Rules are exactly consistent with the record layout or the forms and instructions from which they are extracted. This task should not be done by the person who performed the original extraction.

Initial Validation Checklist

The primary goals of the initial validation process for forms are to ensure that all elements were extracted and that the metadata extracted (e.g., definition, label text etc.) were extracted exactly as they appear on the form and instructions.

1. Verify that each data element in the extracted spreadsheet has a unique ID and unique Data Element Name.
2. On a printed copy of the form:
 - a. Identify each label with an “L” and a number (L1, L2, L3 ...).
 - b. Identify each data field with an “F” and a number.
3. On the form instructions, identify each instruction, definition and business rule with an “I”, “D”, or “B”, respectively, and a number (e.g., I22, D3, B16).

4. Verify that every label and data field identified above appears in the extracted list of data elements. Add any missing data elements.
5. Verify that every relevant instruction, definition and Business Rule appears in the extracted list of data elements. Add any missing instructions, definitions and business rules in their respective columns.
6. Verify that all extracted data element attributes are worded exactly as they appear on the forms and instructions.
7. Verify that all definitions conform to the following criteria (see *Human Services Domain Data Standards Naming Conventions and Controlled Vocabulary*):
 - a. They do not refer back to the data element name as part of the definition.
 - b. They are not instructions on what values should be entered or are valid.
8. Verify that all Business Rules conform to the following criteria:
 - a. The Business Rules must be specific to the data element.
 - b. Define the values to be entered or that are valid.
9. Verify that all check boxes that appear on the form are noted as being check boxes in the Business Rules column.
10. Spell-check the entire list of extracted data elements.
11. Change the date and version number of the extraction spreadsheet.
 - Update the Validated By/Validated Date column for each data element validated.

ID	Data Element Name	Human Services Domain Standard Name	Label	Definition	Authoritative Source	L	T	Value Domain	Business Rules	Issues/ Comments	Mapping	Validated By/ Validation Date
												Validated by: John Doe 5/12/2006

3.3 Identify Anomalies

Before migration of the elements into the Data Standards Registry, and prior to the final validation, anomalies that have been identified must be resolved.

Anomalies come in many forms. Following are examples of data element anomalies and tips for resolving them.

- Inconsistent or Ambiguous Definitions for the Same Data Element. Sometimes the definitions for the same element are not consistent. For example:

The definition for Benefit Amount as provided in the Data Access program’s SWA Output Detail record is:

“This field will contain the gross amount of UI benefits, prior to any deductions, paid to a claimant during the reporting quarter as provided on the UI record submitted to the NDNH.”

While the same data element is defined in the FCR Interface Data Guide as:

“The monetary amount of Unemployment Insurance benefits a person received during a Reporting Period.”

There are two anomalies in the example above: the 1) SWA definition defines “reporting period” as a “reporting quarter” and describes the source for the data, while 2) the FCR definition is less specific regarding the “Reporting Period” and does not specify the source for the data.

- Business Rules as Definitions/Inadequate Definitions. Definitions provided from the source documentation may be inadequate, or the definitions provided are business rules. Example:
 - **Record Identifier**. One source defines it as: “The Record Identifier code defines the type of input record being submitted or the type of output record being sent by the FCR.” This is adequate as a definition.
 - Another source defines a **Record Identifier** as: “This field must contain the characters 'WDCTRL'”. This is not a definition but a Business Rule.
- Inconsistent Format. The value for **Date of Birth** is defined differently in two different sources:
 - **TITLE XVI Date of Birth**: A valid date in CCYYMMDD format, or space.
 - **Date of Birth**: MMDDCCYY (month, day, and year) of the recipient.

Research is necessary to determine whether both formats are valid within their contexts or if they should be identical.

- Use of Abbreviations. DOB and SSN should be expressed as ‘Date of Birth’ and ‘Social Security Number’.

3.4 Integrate and Collate Data Elements

Determine if data elements are similar or identical. Data elements are considered similar if the elements share the same Definition and Authoritative Source but the Business Rule, Length or Type differs.

Data elements that have the same Definition, Authoritative Source, Business Rules, Length and Type can be considered identical. Each instance of a set of identical data elements will be assigned the same Human Services Domain Standard Name, and the Mapping column for each element will contain a reference to the other instances (using the unique ID for those data names). Each data element will retain its unique ID.

A notation in the Issues/Comments column should be made to indicate the similarity. This may require further research to determine if they are identical or not. Each instance of similar data elements will be assigned a Human Services Domain Standard Name, with qualifiers to indicate they are not identical. If, following further research, elements are determined to be identical; their names will be changed accordingly.

It is important to qualify the term ‘same’ as it is used above. In the definitions above, ‘same’ refers to the meaning of the Definition; it does not require that the Definitions match word for word. Type and Length must be a match to be considered ‘same’.

3.5 Final Validation

The extracted data elements must once again be validated by individuals or teams who did not perform the initial harmonization tasks. This ensures a greater degree of objectivity and conformity in the data element definitions.

Final Validation steps:

1. Identify an Authoritative Source for each data element definition following guidelines provided in Section 2.6.
2. Provide all relevant Business Rules for each data element. See Section 2.8 for further information.
3. Assign a Human Services Domain Standard Name. See Section 2.3 for more information.
4. Assign Length and Type Attributes.
5. The final step in the harmonization process is to define a Length and Type to each of the data elements. Record layouts and data dictionaries will typically have already defined element lengths and types, but elements derived from paper forms typically do not.
 - For identical data elements, length and type will be identical.
 - The Length and Type can be derived from other data elements used in other applications by examining the consolidated list of data elements for similar data elements.

For questions or comments about this document, contact the Human Services Domain POC at interoperability@acf.hhs.gov.

Appendix A: Examples

Data Element Name

A “**Date**” on a form, if associated with a signature, could be extracted as **Submitted Date** or **Signed Date**, dependent on how it is described in the instructions and/or its context in a form.

Definition

The following are examples of valid definitions:

- **Obligor:** “The individual who owes or is alleged to owe support. This term includes alleged or putative fathers whose paternity of the child(ren) has not yet been established.”
- **Paternity:** “The legal establishment of fatherhood for a child, either by court determination, administrative process, or voluntary acknowledgment. A paternity acknowledgment involves the legal establishment of fatherhood for a child through a voluntary acknowledgement signed by both parents as part of an in-hospital or other acknowledgement service.”
- **Income-withholding Order:** “An order or other legal process directed to an obligor's employer (or other debtor), as defined by the income-withholding law of this State, to withhold support from the income of the obligor.”

Length

A Social Security Number, for example, may have a 9-byte length and an alphanumeric type.

Business Rule

Business Rules define formatting, possible values and formats for a data element. The following are examples of Business Rules:

- Formatting. “This field will contain the date stamp in CCYYMMDD format from the record submitted” (from the SWA Input Header Record [among others]).
- Reports. “Required for the following input record: FCR Input Case Record with Action Type Code 'A'. Optional for the following input record: FCR Input Case Record with Action Type Code 'C'. Conditional for the following output records: FCR Case Acknowledgement/Error Record.”
- Validation Rules. When rules are used to validate data instances, (e.g. validating a Social Security Number (SSN), returned results provide additional information about valid values, such as:
 - 1 - SSN not assigned by SSA
 - 5 - SSN and Name do not match
 - 6 - SSN invalid (spaces, all zeros, sixes, nines are not valid)
 - 7 - Borrower's Name is missing (must be at least one character in the first name and two characters in the last name)” (DoEd Verification Code from the DoEd Output Error Response Detail Record Layout).

Issues/Comments

Examples of Issues/Comments are:

- **OPTIONAL:** This would be used if the instructions on a form say something like, “Enter information here only if check box above is checked.”
- **OPTIONAL:** The instructions for the form say “if applicable and if known.” This implies the name is optional.
- **CONDITIONAL:** Either this check box is checked or another check box is checked. Both cannot be checked.
- **ANOMALY:** "Responding" is defined differently in other sources.

Authoritative Source

Examples using the name Participant Type are:

- **Participant Type** is defined in the FCR Interface Guidance Document as “The code that defines the participant's relationship in the child support case,” which is a valid definition.
- A statutory source for **Participant Type** is found in the Code of Federal Regulations at 45 Section 307.11, and is defined as, “The custodial party, non-custodial parent, putative father, or child, associated with a case or support order contained in the State or Federal case registry.” This statutory definition, along with its citation, will be entered in the Authoritative Sources column.