Social Needs and Services Structured Data Extract Implementation Guide

February 2025

Version 1.4

Prepared for:
Administration for Community Living

MITRE | Center for Transforming Health

Operator of the CMS Alliance to Modernize Healthcare (The Health FFRDC)

This and other CODI resources are available at https://mitre.github.io/CODI/.

Notice

Authors

The MITRE Corporation, McLean, VA

Record of Changes

Version	Date	Author/ Owner	Description of Change
1.0	July 2024	E Kraus; J Irving; K Hay	First version of implementation guide
1.1	August 2024	E Kraus	Clarification of PRESCRIBED_SESSION_FREQUENCY. Directions for generating identifiers. Formatting.
1.3	December 2024	E Kraus; K Hay; D Hall	Deprecated PRO_CM and added QUESTION and QUESTION_RESPONSE tables Added PROGRAMID to ASSET_ENROLLMENT and ASSET_DELIVERY Removed PROGRAM_ENROLLMENT_ID from ASSET_DELIVERY Added ASSET NAME and ASSET_DESCRIPTION to ASSET_ENROLLMENT Added REFERRALID to PROGRAM_ENROLLMENT Added PROGRAM_ENROLLMENT_ID to SESSION Consolidated multiple different code system value sets into one CODE_SYS value set. General clarifications and formatting. Updated data model name from ACL CODI Data Model to CODI Data Model to align with branding. Replaced ACDM with CODI Data Model throughout.
1.4	February 2025	E Kraus; K Hay	Minor modifications to DEMOGRAPHIC table. Updated title page. Addition of PRS statement. Addition of GitHub links.

Table of Contents

List of Figures

List of Tables

1 Introduction

Community-based organizations (CBOs) play an essential role in addressing individual health related social needs (HRSN) by providing support services such as home-delivered meals and shelf-stable food to individuals with food insecurity. However, measuring the impact of CBO provided services on meaningful clinical and quality of life outcomes is a challenge because data about service provision often resides in siloed information systems, disconnected from clinical data. Older adults experience social needs at a higher rate compared to individuals less than 60 years old; therefore, the importance of identifying older adults with social needs and connecting those individuals with the right services is elevated. Harmonizing clinical and community data is required to demonstrate the crucial role of CBOs in older adult health.

The Administration for Community Living (ACL) has selected the Clinical and Community Data Initiative (CODI) model[1] to bring together community data from CBOs with clinical data from healthcare providers to improve older adult health. This work is part of an Administration for Community Living (ACL) sponsored project, the Community and Clinical Data Initiative (CODI) Maryland pilot.[2]

The CODI Maryland pilot will combine individual-level social needs and services data from CBOs (community data) with clinical data to evaluate the impact of CBO services on older adult health. As part of the pilot, the Health FFRDC created the CODI Data Model to specify a common data format and structure for extracting, normalizing, and sharing social needs and services data.

This document describes the CODI Data Model and provides guidance for CBOs implementing the CODI Data Model and for organizations receiving and interpreting data received from CBOs.

1.1 Background

The CODI model was pioneered by the Centers for Disease Control and Prevention (CDC) to harmonize clinical and community data for public health research.[3] CDC supported the development of the CODI model and two CODI implementations in Colorado (for childhood obesity research) and in North Carolina (for chronic disease research).

During the development of the CODI model, the Health FFRDC created two complementary common data models: the research data model (RDM) and the record-linkage data model (RLDM). The RDM stores clinical and social service data. The RLDM is used only for implementing privacy-preserving record linkage (PPRL). A detailed description of the CDC CODI data models is available in 6.1Appendix A.

During the CODI Maryland pilot, individual-level clinical and community data will be linked by identifiers, negating the need for PPRL and the RLDM. Of the two CODI data models, only the RDM is applicable to the CODI Maryland pilot work and is the origin of the CODI Data Model.

[1] CODI Toolbox: https://phii.org/course/codi-toolbox/

[2] This and other resources available at https://mitre.github.io/CODI/

[3] CODI is a model for harmonizing clinical and community data, developed by the Centers for Disease Control and Prevention; CODI home page: https://www.cdc.gov/obesity/hcp/codi/index.html

1.2 Purpose

The purpose of this document is to describe CODI Data Model and provide guidance to implementers to build and populate relevant CODI Data Model data tables with data from their source information systems. This document provides:

- Descriptions of CODI Data Model and the structured data extract (SDE)—a subset of CODI Data Model tables
- General expectations regarding how to implement SDE
- Specific guidance for each SDE table and their individual data attributes

A CODI Data Model Data Dictionary[4] (DD) workbook developed in conjunction with this guide provides the definitions for CODI Data Model tables and attributes as well as defined value sets that apply to certain attributes. The DD workbook is useful to both CODI Data Model implementers and end-users for understanding how to transform data from source systems into the CODI Data Model and how to interpret data received in the CODI Data Model format and structure. Notably, the table and attribute descriptions in this document exclude some technical detail provided in the DD for brevity.

Some of the content in this guide and the DD workbook has been adopted or adapted from other sources, including previous CODI materials or from the Patient Centered Outcomes Research Network (PCORnet) common data model; the official source of record for content from other sources should be considered authoritative.

1.3 Scope

This document provides general guidance for organizations implementing the CODI Data Model and SDE to extract and share social needs and service data with another entity.

For certain topics that are specific to each organization, this implementation guidance will not provide sufficient detail. Implementers are expected to adhere to the spirit and guidance provided in the implementation guide during their CODI Data Model and SDE work—supplementing the implementation guide with their own institutional knowledge and implementation specifications unique to their information systems, when appropriate.

1.4 Audience

The primary audience for this document is CODI Data Model implementers who are presumed to be CBOs that screen and assess individuals for social needs, enroll individuals in social care services, or provide social services directly to individuals. Examples of implementers are Area Agencies on Aging (AAAs), community care hubs, and Meals on Wheels organizations.

Secondary audiences include organizations holding clinical data and/or receiving social needs and service data from implementers such as health information exchange (HIE) organizations, healthcare organizations, and government organizations who can use this implementation guide to interpret and process social needs and service data received in the CODI Data Model format.

Stakeholders in communities outside the CODI Maryland pilot jurisdiction may be interested in this implementation guide to consider the feasibility of implementing a similar data sharing effort.

[4] CODI Data Model Data Dictionary Link: https://mitre.github.io/CODI/

Future jurisdictions may be interested in this implementation guide to understand the work of CODI Data Model implementation.

2 CODI Maryland Pilot Overview

This section provides a brief overview of the CODI Maryland pilot design, defines the organizational roles that are relevant to CODI Data Model implementation, and a describes the CODI Data Model and SDE.

2.1 CODI Maryland Pilot Design

The CODI Maryland pilot design is an approach for connecting clinical and community data centered around HIEs. The rationale behind an HIE-centered model is to leverage existing technical infrastructure provided by the HIE, rather than replicating or duplicating technology, knowledge, and expertise that HIEs already possess. CBOs collect many different types of data due to the diversity of social needs and services provided; CODI Data Model provides the standard format for social needs and service data extraction, transformation, and exchange required for linkage to clinical data.

Figure 1 illustrates the proposed data flow where direct social service providers extract social needs and service data from their information systems, transform that data into a standard format (e.g., CODI Data Model), and securely share that data with the HIE. The HIE then processes those data; links individuals, social needs, and service data with clinical data; and uses linked data to compute outcomes.

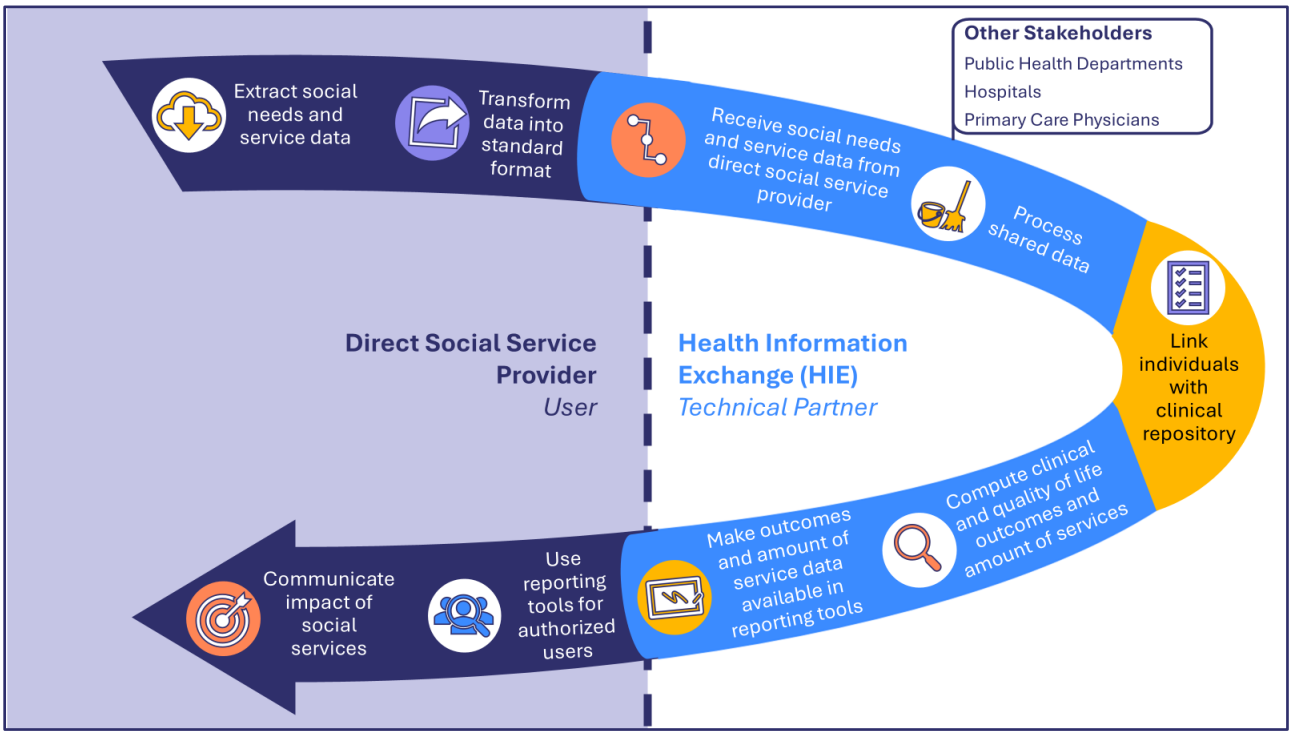

Figure 1: CODI Maryland Pilot Design

2.2 CODI Maryland Pilot Operational Concept and Roles

The following two operational roles are required for CODI Data Model implementation:

- Direct social service provider (e.g., CBO)
    - Delivers services and documents social needs and service data in one or multiple information systems
    - Implements CODI Data Model to normalize social needs and service data to a common format
- HIE
    - Receives social needs and service data in the CODI Data Model structure and format
    - Links data received from CBOs with clinical data received from healthcare provider organizations.

2.3 Overview of CODI Data Model

CODI Data Model contains tables and attributes to store data about individuals, social needs, social services, programs, clinical data, and cost needed to measure the impact of social services on clinical and quality of life outcomes (See CODI Data Model Data Dictionary[5]). Services is a broad category that include assets (e.g., rides, housing vouchers, meals) and structured programs (e.g., diabetes self-management program) (see Section 3.4).

Table 1 summarizes the major conceptual components of the CODI Data Model.

Table 1: Conceptual Components of the CODI Data Model

Concept	CODI Data Model Tables	Description
The individual	DEMOGRAPHIC PRIVATE_ADDRESS	Includes demographic information about a patient or program participant; includes personally identifiable information (e.g., name and address) to link individual health records.
Social needs	DIAGNOSIS QUESTION QUESTION_RESPONSE	Includes diagnostic codes for social needs, responses to social needs screening questionnaires, and social need-specific assessments such as a food insecurity assessment.
Referrals between organizations	ORGANIZATION PROVIDER REFERRAL	Includes incoming and outgoing electronic referrals within and across organizations for clinical needs (e.g., a specialist) or social needs such as a referral to be screened for social needs or a referral for specific services to address an identified need. Also includes reference information about organizations sending or receiving referrals.

[5] CODI Data Model Data Dictionary: https://mitre.github.io/CODI/

Concept	CODI Data Model Tables	Description
Enrollment and delivery of assets intended to address an identified need	ASSET_DELIVERY ASSET_ENROLLMENT ORGANIZATION	Includes enrollment in service delivery (conceptualized as an asset) such as home delivered meals to address food insecurity and delivery records for those assets. Also includes reference information about organizations enrolling in or delivering assets and programs that the assets are associated with (if applicable).
Enrollment and delivery of programs (e.g., evidence-based programs)	CURRICULUM_ COMPONENT ORGANIZATION PROGRAM PROGRAM_ENROLLMENT SESSION	Includes details about program or intervention aims and settings (PROGRAM); how the interventions are structured (CURRICULUM COMPONENT); who is enrolled in programs and who is administering the intervention and how (PROGRAM_ENROLLMENT, SESSION); and if a program (e.g., Diabetes Prevention Program) was delivered (SESSION). Also includes reference information about organizations enrolling in or delivering the program.
Clinical care	CONDITION ENCOUNTER DIAGNOSIS LAB_RESULT_CM PRESCRIBING PROCEDURE PROVIDER REFERRAL VITAL	Includes information about an individual's interactions with the healthcare delivery system.
Cost of care	COST	Includes information about the amounts charged.

CODI Data Model is designed as a relational database that stores and provides access to data points that, when combined and related to one another, represents information about an individual. All the information about one person is stored across multiple data records in multiple tables and is linked with a common record key value. See 6.1Appendix B for more information about relational databases and keys.

2.4 Structured Data Extract

The SDE is a subset of the CODI Data Model and includes the 12 CODI Data Model tables that capture social needs and services data. The SDE is designed to support the harmonization of data across a wide range of organizations that address different social needs and deliver a range of social services. The SDE tables are:

1. DEMOGRAPHIC
2. PRIVATE_ADDRESS_HISTORY
3. QUESTION
4. QUESTION_RESPONSE
5. ASSET_ENROLLMENT
6. ASSET_DELIVERY

7. PROGRAM
8. PROGRAM_ENROLLMENT
9. SESSION
10. CURRICULUM_COMPONENT
11. ORGANIZATION
12. REFERRAL

Figure 2 displays the 12 SDE tables and relationships between tables that make up a person's entire social needs and services story. Guidance for each table is provided in Section 4. The SDE is centered around the person and includes mostly person-level or individual-level data. The DEMOGRAPHIC table is the central table for the person record. The DEMOGRAPHIC and PRIVATE_ADDRESS_HISTORY tables hold the contact information that will be used to link individuals across organizations. Social needs data are stored primarily in QUESTION and QUESTION_RESPONSE. Social services data are stored either as an asset in ASSET_ENROLLMENT and ASSET_DELIVERY or as a program in PROGRAM, PROGRAM_ENROLLMENT, SESSSION, and CURRICULUM_COMPONENT. The ORGANIZATION table is a reference table to describe the organizations enrolling individuals in programs or assets, or delivering assets or program sessions (See Appendix B for more information on reference tables). The REFERRAL table stores electronic referrals sent or received for social needs or services represented in other tables.

Each SDE implementation may not require the use of all SDE tables; however, the DEMOGRAPHIC, PRIVATE_ADDRESS_HISTORY, and ORGANIZATION tables are always required.

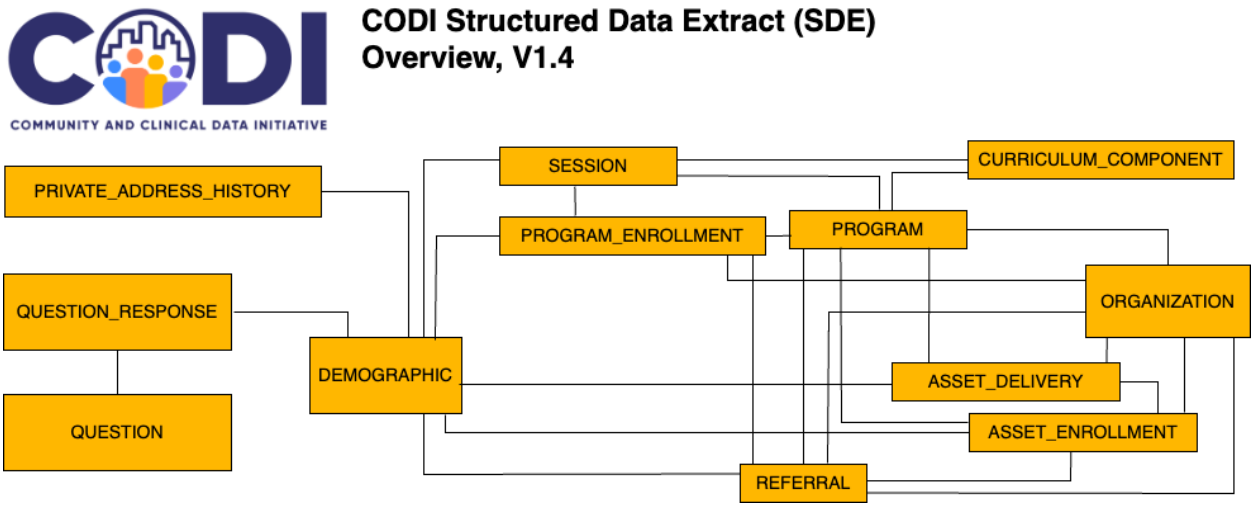

Figure 2: Structured Data Extract Table Relationship Diagram

3 Implementation Guidance

This section provides guidance to an implementer tasked with creating the SDE to represent the work of their organization. This guide assumes that SDE implementers:

- are CBOs, AAAs, or community care hubs that offer programs and services.
- are familiar with their organization's social needs screening activities and social services, corresponding data collection practices, and the format of that data stored in their information system (source information system).
- will determine which SDE tables to build and populate based on the services an organization provides (see Table 2).

- are familiar with foundational data engineering concepts (Appendix B provides additional explanations of key topics).
- will have different resulting data, documentation practices, and information systems making each SDE implementation unique.

3.1 Approach

A recommended approach to implementation includes the following seven activities. More detailed guidance, when available, is referenced in each activity.

1. **Discovery:** Review resources to identify the initial scope of data for SDE inclusion.
    - Review this document and the CODI Data Model Data Dictionary to understand the contents of social needs and service data.
    - Determine what data from each source information system will be included in the SDE implementation.
    - Determine which SDE tables will be built and populated based on implementing organization's work (See Section 3.4)
2. **Extract Preparation**: There are several steps in preparing a data extract. The ability to configure the extract at each of these steps will vary by information system. Some systems may not provide options for selecting or filtering data, requiring extra work, or transformation, to create the SDE. The steps include:
    - Identify what source information systems data will be extracted from.
    - Determine the population of interest by defining what individuals will be included and excluded.
    - Establish the time period for the extract by selecting how far back the data extract will extend.
    - Identify which data elements from the information system will be in the extract.
    - Define an approach for how data will be extracted and transformed. This is described in Activities 4 and 5 below.
    - Establish the frequency of the data extract process.
3. **Service Modeling**: For the services that an organization wishes to include in the SDE, determine if each service will be modeled as a program or as an asset. See Section 3.5 for more information.
4. **Attribute and Value Mapping**: From each attribute that will be included in the SDE, map the attribute name from the source information system to the SDE table and attribute name and map response values from source information systems to SDE allowed values and format. See Section 3.6.3 for more information about value sets. An SDE Mapping Worksheet is an available CODI resource to support implementers in conducting and recording mappings.
5. **Data Transformation**: Use mappings to establish and test a process that extracts, transforms, and loads (ETL) data from source information system into an SDE. Transformation may also require that some information is purposefully discarded as a part of the process. For example, some information systems may contain test records that are not associated with real people. These records should be omitted when transforming data into the SDE. See Section 3.6 for more information. Key steps include:
    - Establish ETL process

- Test ETL with real-world data and validate that resulting SDE meets specifications.
- Consider how new values in the source information that require mapping updates will be identified and updated accordingly in ETL.

6. **Pilot SDE**: Complete a full ETL run to generate an SDE from source information systems. If applicable, test transmission of SDE to a desired recipient organization using secure methods and confirm receipt. See Section 5.1 for more information.
7. **Deploy SDE**: When possible, transition from manually running the ETL process to a regularly scheduled, automated run of the ETL process. CODI recommends that implementers share data at least monthly. See Section 5.2 for more information.

3.2 Implementation Best Practices

Implementers should keep in mind the following best practices:

- **Populate SDE with electronic data only:** It's crucial to maintain the integrity and accuracy of data by only inputting electronic data into the SDE. Avoid manually entering or reverse-engineering data, as this can lead to errors and inconsistencies and can be difficult to sustain.
- **Avoid hard coding unrecorded data:** Do not hard code data that isn't captured in your information system. Do not impute or derive if not expressly defined in the source system. This practice can lead to incorrect assumptions and inaccurate data representation. Instead, work on improving your data capture methods to include all necessary information.
- **Be intentional about mapping:** When mapping data, be deliberate about how you handle categories for null values like "no information," "unknown," and "other." These categories can often be overlooked or mishandled, leading to a loss of valuable data insights. Develop a consistent strategy for dealing with these types of data. See Section 3.6.5 for guidance on when to use different types of null values.
- **Use fields for multiple attributes:** It's acceptable to use one field in an information system to map to multiple SDE attributes. This can happen when the source data system combines more than one concept into a single field. A common example to visualize this would be where blood pressure is collected as one value called "Blood Pressure" rather than distinct fields for systolic and diastolic measurements. However, ensure that this doesn't lead to confusion or misinterpretation of the data. Always clearly document your mapping decisions to ensure they can be understood by others. The SDE Mapping Worksheet column for Business Logic or Transformation Notes that enables documentation of these decisions.
- **Handle Personally Identifiable Information with care:** Personally Identifiable Information (PII) is any information that can be used to identify an individual. PII attributes in the SDE include name, address, phone number, email and date of birth. Implementers should review and adhere to their organization's policies regarding PII privacy and data protection (See 0 for privacy best practices).

3.3 Choosing a Data Start Date

Implementers should establish a start date for their SDE data extraction. The start date represents the earliest possible date for which implementers will provide social needs and service data. Each implementer will determine a start date for each source information system

they are extracting data from. Implementers are encouraged to collaborate with other CODI stakeholders to align the timing of start dates. For instance, implementers should coordinate with their HIE partner to understand the earliest date on which clinical data is available for linking.

3.4 Determining Relevant Tables for Implementation

Implementers can use the SDE to share data about: individuals, individual-level social needs, enrollment in and receipt of social services, enrollment and participation in programs (e.g., social care or evidence-based programs), referrals between organizations, and organizations providing services, delivering programs, and screening for social needs. However, not all implementers will be engaged in every type of work to generate all different types of data. Instead, implementers will have to determine which SDE tables they need to implement.

Table 2 guides implementers in determining which SDE tables are relevant to their data and necessary to implement.

Table 2. Structured Data Extract Tables to Implement Based on Data Type

SDE Tables	Data Type			
	Program Enrollment and/or Delivery	Asset Enrollment and/or Delivery	Referrals (incoming or outgoing)	Social Needs
DEMOGRAPHIC	X	X	X	X
PRIVATE_ADDRESS_HISTORY	X	X	X	X
ORGANIZATION	X	X	X	
QUESTION				X
QUESTION_RESPONSE				X
PROGRAM	X			
PROGRAM_ENROLLMENT	X			
SESSION	X			
CURRICULUM_COMPONENT	X			
ASSET_ENROLLMENT		X		
ASSET_DELIVERY		X		
REFERRAL			X	
Total	7	5	4	4

Implementers can reference the CODI Data Model Data Dictionary workbook to determine which attributes in the SDE tables that they are implementing are required.

3.5 Modeling Services as Assets or Programs

The SDE provides two options for modeling services: a service may be an asset or a program (See Table 3). Implementers should consider the properties of each service in the context of the

properties of an asset and a program and select the most appropriate approach for each of their services. Implementers are permitted to implement some services as assets and other services as programs.

Table 3: Comparison of Program and Asset Properties

	Asset	Program
Description	• An asset is a resource transferred by an organization to an individual (i.e., meals, vouchers, or rides) • An asset includes data from enrollment and delivery • Assets allow implementers to document only enrollment and delivery and describe the assets as a type and subtype. Thus, if a service consistently includes more components (e.g., education, screening) than delivery of an asset, the service may be better modeled as a program	• A program is a defined sequence of individual or group engagements with a specified aim and structure • Programs may have one or more aims (e.g., nutrition and physical activity). • A program includes date from enrollment and delivery • Programs generally include regularly scheduled sessions that are held for a defined period of time at a specific location • The PROGRAM table describes the location, duration, frequency, and aim of the program • Program curriculum can be described in CURRICULUM_COMPONENT table • Programs may include delivery of assets, such as a diabetes prevention program that includes a meal delivery component
Examples	• Home delivered meals • Supplemental nutrition assistance program	• Fall prevention program • PEARLS depression program

Recipients of the SDE (e.g., HIEs) may want to consolidate assets and programs into one set of tables. The data dictionary includes an example mapping of asset and program tables (See Program Asset Mapping tab in DD).

3.6 Data Transformation Guidance

3.6.1 Allowed Attribute Values

SDE includes attributes of various types. Understanding the types of data attributes can inform an implementer's understanding of what data can be populated in that attribute. Table 4 describes the possible types of attributes.

Table 4: Description of Attribute Types

Attribute Type	Description	Data Dictionary Column: Value(s)
Coded or Categorical	An attribute with an associated value set of categorical or coded values listed in the CODI Data Model Data Dictionary CodeSet Dictionary tab.	UML Type: [CODESET_NAME], CHAR (X), VARCHAR(X), YES/NO
Uncoded or unstructured	An attribute that is a character or string data type and may be populated with text of a defined length (number of characters). This attribute can hold numbers but is formatted as text.	UML Type: string, CHAR (X), VARCHAR(X)
Numeric	An attribute defined as an integer, decimal, real, or float data type that may be populated only with a number.	UML Type: float, integer, NUMERIC(X), Boolean
Date field	An attribute defined as a date data type using the date types available in the database management system. Dates should be formatted as MMDDYY. If implementers have date times, trim the time off when loading those values into a date field.	UML Type: Date
Time field	An attribute defined as a time data type using the time types available in the database management system. Time does not include the day.	UML Type: Time
DateTime field	An attribute defined as a datetime data type using the datetime types available in the database management system.	UML Type: dateTime
Required	An attribute that shall not be without a value. This attribute cannot be null and must be populated with a value.	Requirement: Required
Optional	An attribute that exists in SDE but is allowed to be without any value. This attribute is only populated when data exist in the source information system.	Requirement: Optional

3.6.2 Generating Identifiers

Unique Identifiers (IDs) are an essential component of SDE tables and intended functionality. Unique means that each value is different from all other values in that attribute so that no two rows have the same identifier value. Unique identifiers are used to establish and enforce relationships between tables (referential integrity). For example, a foreign key in one table can reference a primary key in another table, creating a link between the two tables.

SDE relies on a unique identifier for each person (PATID) that is the primary key for the DEMOGRAPHIC table and foreign key in many other tables. For other SDE tables, a unique identifier serves as a primary key to uniquely identify each row (e.g., ASSET_ENROLLMENT_ID and ASSET_DELIVERY_ID).

To generate unique identifiers, implementers may use existing IDs or generate their own identifiers. Implementers are cautioned that existing identifiers should only be used if permitted and if existing identifiers pose a very small risk for reidentification. The Health Insurance Privacy and Portability Act provides a deidentification standard that can be used to assess if identifiers are deidentified.[6] Identifiers must be unique, meaning they cannot be reused and must not be NULL. Implementers may use numbers or characters to generate identifiers and are discouraged from using pieces of personally identifiable information (e.g., the first two letters of a person's last name) to prevent risk of reidentification. Unique identifiers should be persistent over time meaning that one person has the same PATID across SDE refreshes.

If the organization that will receive the SDE (e.g., HIE) is already acquainted with a unique identifier from the implementing organization, reusing the existing unique identifier that the recipient is familiar with may be advisable. If the implementing organization does not have a master identifier for its services, sending non-unique IDs may impact linkage when sharing the SDE with the receiving organization. Implementing organizations should discuss how to handle this situation with the receiving organization to mitigate potential linking issues, if applicable.

3.6.3 Keys and Referential Integrity

Data across SDE tables are related through data attributes referred to as keys. Keys allow users of the data to combine and merge data about a single individual into a longitudinal record. For example, keys allow a social needs screening question and response to be connected to a distinct individual. In this case, the PATID field in the QUESTION_RESPONSE table is a foreign key, referring to the PATID in the DEMOGRAPHIC table to establish the relationship.

Referential integrity is a property of data which, when satisfied, requires every value of one attribute (column) of a relation (table) to exist as a value of another attribute in a different (or the same) relation. In simpler terms, it ensures that relationships between tables in a database remain consistent. This is typically enforced through the use of foreign keys. The SDE includes many keys to ensure referential integrity throughout the database so that keys allow pieces of data to be referenced completely and consistently.

3.6.4 Mapping Value Sets

Mapping means creating associations between two distinct data models to convert data from one format to another, often from a local value set to a standardized one, such as SDE.

It is expected that implementers will map many local values to a SDE value set. Mapping is a necessary step to ensure compatibility and interoperability. Implementers are responsible for this task, and they should convert the values from their source information systems to the defined SDE value sets.

For example, if a local value set uses the term "Female" to denote female sex, and the SDE value set uses the term "F," the implementers would map "Female" from the local value set to "F" in the SDE value set. Mapping ensures that the data can be understood, used correctly, and retains the meaning it had in the source information system.

[6] HIPAA De-identification Standard: https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html#standard

3.6.5 Missing Data

During the mapping process, implementers will encounter SDE attributes or individual records where data from their source information system are missing. For each attribute, data may be always missing or frequently, for example if an individual's middle name is not collected or if race is not collected for 10 percent of the individuals that they see.

SDE uses the Health Level 7 International (HL7) conventions[7] to guide implementers about what to do when data is unavailable, missing, or unknown in their source information system:

- For an **optional attribute** that is not present in the source information system (e.g., not collected or collected but not documented), regardless of the data type, implementers should generate or use a null value.
- For **required attributes**:
  - A coded text attribute that is present in the source system and is null or blank, SDE prefers NI (no information).
  - A coded text attribute whose source value is an explicit unknown value, SDE prefers UN (unknown).
  - A coded text attribute with a code value that cannot be mapped to an SDE allowed value, SDE prefers OT (other). For some attributes, a qualified other option may be available (like OS for Other Service Provider).

3.6.6 Data Standards and Value Sets

SDE uses data standards to standardize the same type of data into a consistent format, set of values, and unit for some attributes. Value sets are a subset of concepts (each concept represented by a specific code) either developed for this project or drawn from existing code systems or standards (e.g., postal codes, disease codes). The benefit of using a value set rather than free text data in certain attributes is so analysts can compare and combine attribute values from different organizations' datasets to draw conclusions about social needs, services, and health impacts. The following data standards are used in SDE to codify data about social needs and services.

- **Gravity Project (Gravity)**[8]: Gravity is a leading data standard for social needs and services data. SDE uses the Gravity domains to identify the social need that a referral is for or an asset is addressing. More information on Gravity is in 0.
- **Logical Observation Identifiers, Names, and Codes (LOINC®)**[9]: LOINC is a common language (set of identifiers, names, and codes) for identifying health measurements, observations, and documents. LOINC is most often used to standardize lab tests and results but also provides codes for many other types of health information. SDE uses LOINC codes to identify questionnaires that collect patient reported information (like social needs screeners) and the individual questions and responses in those instruments. Gravity has mapped questions from common social needs screening instruments to LOINC codes and published those mappings as value sets to the Value Set Authority Center (VSAC).[10]

[7] HL7 guidance on unavailable data: https://terminology.hl7.org/1.0.0/CodeSystem-v3-NullFlavor.html
[8] The Gravity Project: https://thegravityproject.net/
[9] LOINC: https://loinc.org/get-started/what-loinc-is/
[10] Value Set Authority Center: https://vsac.nlm.nih.gov/

- **Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT):** SNOMED is a comprehensive, precise, and multilingual health terminology used across the world. SNOMED CT is developed collaboratively to ensure it meets the diverse needs and expectations of the worldwide medical profession and assists with the electronic exchange of clinical health information. SNOMED CT can be mapped to other coding systems, such as International Clinical Diagnosis (ICD)-9 and ICD-10.[11] SDE uses SNOMED CT procedure codes to represent delivery of an asset (using procedure code value sets recommended by Gravity). Gravity has recommended that the delivery of social services be conceptualized and modeled as a procedure, synonymous with a medical procedure.

Refer to the CODI Data Dictionary Workbook[12] Codeset Dictionary tab for a complete listing of all value sets.

3.7 Data Cleaning

In general, the SDE should be populated with structured data extracted from a source information system, unless specified otherwise. Implementers are encouraged to clean SDE data with any existing data cleaning processes or tools native to their data systems and governance requirements and are not expected to build any new data cleaning processes as part of the SDE implementation.

Example data cleaning strategies are as follows:

- **Removing Duplicates**: Duplicates can occur due to various reasons such as data entry errors or merging of data from different sources. Implementers are encouraged to remove duplicates.
- **Removing White Space**: Leading spaces or blanks can obscure a numeric or character entry. Implementers are encouraged to trim extra blanks or spaces from numeric and character values.
- **Handling Missing Values**: Missing data can lead to inaccurate analysis. Implementers are encouraged to examine the frequency of missing data with an emphasis on attributes that are important for analysis and either remove the rows with missing data or fill them with appropriate values.
- **Outlier Detection**: Outliers can significantly affect the results of analysis. Implementers are encouraged to examine data for outliers such as delivery of 200 meals to an individual in one day.
- **Correcting Inconsistent Values**: Implementers are encouraged to identify inconsistent values where the meaning is the same but the spelling or presentation is different. One example is making sure state entries are always spelled out and that no two letter state abbreviations are included.
- **Formatting Data**: This involves converting the data into a format that is suitable for analysis. For example, you might need to convert a string field into a numeric field.

[11] SNOMED: https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.html
[12] CODI Data Model Data Dictionary: https://mitre.github.io/CODI/

- **Removing Unnecessary Data**: This involves identifying and removing data that is not needed or accurate. For example, implementers should ensure that test data is not included.
- **Converting Categorical Data**: This involves converting categorical data into a format that can be used for analysis. For example, implementers may need to convert a categorical field into a binary field.

3.8 Protecting Personally Identifiable Information

This document assumes that implementers may handle and potentially share PII with an external organization. PII is sensitive as it can be used to identify individuals receiving clinical care or services.

Implementers are expected to remain compliant with applicable federal regulation and state statutes regarding the protection and sharing of PII including, but not limited to, Health Insurance Portability and Accountability Act during their activities that relate to CODI.

Implementers should review organizational policies and agreements that specify how and when data can be shared, used, and protected with a focus on the responsibilities of parties involved in handling PII and measures for data security, privacy, and confidentiality.

Implementers should use secure methods for data transmission and ensure that data is stored in a secure environment with access controls in place to prevent unauthorized access.

Implementers are expected to have systems and processes in place for data management, including secure data storage and disposal.

Finally, implementers should have procedures in place for identifying and managing data breaches, including identifying and reporting breaches, as well as measures to mitigate the impact of any breaches. Regular audits and reviews should be conducted to ensure ongoing compliance with data protection regulations and standards. Lastly, training should be provided to staff handling PII to ensure they understand their responsibilities under data protection laws and regulations and are aware of the procedures to follow to ensure data security and privacy.

3.9 Data Validation

Data validation is the process of checking and verifying that the data entered in a system or database is accurate, complete, and meets specified criteria or standards. It involves various techniques and processes to clean, correct, and ensure consistency in data, thereby enhancing its quality and reliability for further processing or decision making.

The purpose of data validation is to ensure the accuracy, quality, and reliability of data by checking if it meets the specified format, rules, and defined criteria. Data validation helps in identifying and correcting errors, inconsistencies, and discrepancies in data, thereby preventing the use of incorrect or misleading information in decision-making processes. Data validation also aids in maintaining the integrity of the data, ensuring that it is complete and consistent, which is crucial for accurate analysis, reporting, and forecasting.

As feasible, implementers are encouraged to validate their data extract through the following strategies:

- Cross-check a random sample of the information in the SDC with the original source information system or data source to ensure accuracy and consistency in the details such as the number of meals delivered, addresses, and delivery times.

- Check for missing or incomplete entries, discrepancies, or anomalies that could indicate errors or inaccuracies in the data.
- Confirm that the data aligns with the organization's operational records, such as delivery logs or customer feedback, to ensure the data accurately reflects the organization's service delivery.

4 Describing Structured Data Extract Tables and Attributes

This section describes the tables and attributes in the SDE and provides guidance on which data from the organization's information systems should go in those structures; table and attribute names appear in ALL CAPS. It also explains the specific intent for the tables and attributes, and rules for the attribute values captured.

Information on requirements, value sets, and data types are not included in the table summaries below; for a comprehensive description of all tables, attributes, value sets, and other requirements, see the accompanying CODI Data Dictionary Workbook.

Guidance is provided for SDE tables in the following order: DEMOGRAPHIC, PRIVATE_ADDRESS_HISTORY, QUESTION, QUESTION_RESPONSE, ASSET_ENROLLMENT, ASSET_DELIVERY, PROGRAM, PROGRAM_ENROLLMENT, SESSION, CURRICULUM_COMPONENT, ORGANIZATION, and REFERRAL.

4.1 DEMOGRAPHIC

The DEMOGRAPHIC table contains one record for an individual person. Implementers should create a DEMOGRAPHIC record for individuals with any screening responses (questionnaire responses), enrollment in a program or assets, receiving assets, or attending any sessions (as part of a program). Table 5 provides implementation guidance for the DEMOGRAPHIC table.

To preserve referential integrity, there must be a DEMOGRAPHIC record (a PATID) for any individual for whom information exists in QUESTION_RESPONSE, ASSET_DELIVERY, ASSET_ENROLLMENT, PROGRAM_ENROLLMENT, SESSION, or REFERRAL. Conversely, every PATID in DEMOGRAPHIC should have corresponding record(s) in at least one other SDE table. Individuals without records in other SDE tables are omitted because adding individuals to the DEMOGRAPHIC table without other information about social needs and services introduces an unwarranted risk. For example, an individual who has no program participation information, no other self-reported health or social status information, and is enrolled in or receives no assets, should not be included in the DEMOGRAPHIC table.

Individuals represented in the DEMOGRAPHIC table must be unique; duplicate records for the same individual can create collisions when using this data. Implementers are encouraged to remove duplicates for the same individual by merging duplicate records or selecting one best record for the individual.

The DEMOGRAPHIC table includes individual names, phone numbers, email, and birth date for use to link social needs and service data for the same individual across different data sources. Names, phone numbers, email, and birth date are PII and should only be shared with trusted sources in compliance with applicable rules, regulations, and data use agreements through secure means. When multiple values exist for DEMOGRAPHIC attributes, implementers should select the one best value for the person's DEMOGRAPHIC record.

The DEMOGRAPHIC table includes attributes for race and ethnicity (i.e., RACE and HISPANIC). Implementers may have questions about how to implement these attributes and which values to export from the organization's source information system. Race and ethnicity are similar but distinct concepts that are ideally collected separately using recommended value sets. Many organizations do not currently collect these attributes separately or may use values not in the recommended value sets. The DEMOGRAPHIC table provides value sets RACE and HISPANIC attributes that are adopted from the recommended value sets.

Implementers are encouraged to extract all relevant race and ethnicity data that exists in their source information systems and map those data to the SDE attributes RACE and HISPANIC where feasible and with the highest level of accuracy. The implementer may need to transform the source codes or text for these attributes to those codes specified in SDE value sets. Implementers are discouraged from conflating distinct concepts or imputing values.

Table 5: Descriptions for DEMOGRAPHIC Attributes

Attribute	Description	Example
PATID	Arbitrary person-level identifier. Used to link across tables.	GREP9892
PAT_FIRSTNAME	Given name of the patient.	PATTY
PAT_MIDDLENAME	Middle name of the patient.	
PAT_LASTNAME	Surname or family name of the patient.	GREEN
PAT_MAIDENNAME	Surname or family name of the patient prior to marriage.	
PRIMARY_EMAIL	Primary e-mail address for the patient.	pgreen@mail.com
PRIMARY_PHONE	Primary phone number for the patient (if known). 10-digit U.S. phone number.	555-867-5309
PRIMARY_PHONE_TYPE	Primary phone type for the patient (if known) (e.g., home, cell, work). 10-digit U.S. phone number.	M
SECONDARY_PHONE	Secondary phone number for the patient (if known). 10-digit U.S. phone number.	
SECONDARY_PHONE_TYPE	Secondary phone type for the patient (if known) (e.g., home, cell, work). 10-digit U.S. phone number.	
BIRTH_DATE	Patient's date of birth.	12/09/1949
SEX	An individual's biological classification.	F
RACE	Please use only one race value per patient or enter 06 for Multiple race.	03
HISPANIC	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.	N
PAT_PREF_LANGUAGE_SPOKEN	Preferred spoken language of communication as expressed by the patient.	ENG

4.2 PRIVATE_ADDRESS_HISTORY

The PRIVATE_ADDRESS_HISTORY table contains one record per each address associated with an individual. When addresses are available, implementers should generate one or more PRIVATE_ADDRESS_HISTORY records for every PATID in the DEMOGRAPHIC table. The record linkage process requires only one address per individual, but additional addresses may improve linkage accuracy. Therefore, implementers should provide multiple addresses when available. Table 6 provides implementation guidance for the PRIVATE_ADDRESS_HISTORY table. Addresses are used to link data records that pertain to the same individual but were extracted from different source information systems. Addresses are also used for presenting data on a map.

Address data are prone to data quality issues. When possible, implementers should ensure that the PRIVATE_ADDRESS_HISTORY records are the highest quality possible by removing spelling errors and invoking address validation tools to correct address errors.

Like DEMOGRAPHIC, the PRIVATE_ADDRESS_HISTORY holds PII and those data should only be shared with trusted sources in compliance with applicable rules, regulations, and data use agreements and through secure means.

Multiple Addresses

Implementers should confirm whether the recipient organization can use historical addresses; if not, implementers should populate only one row with the current address for each individual. In cases where the PRIVATE_ADDRESS_HISTORY contains multiple addresses per individual, the following rules apply:

- Each record must have an ADDRESS_PERIOD_START value and, except for the record of the individual's current address, each record must have an ADDRESS_PERIOD_END value.
- The implementer should indicate the preferred address (where ADDRESS_PREFERRED is set to "Y"). Ideally, the preferred address will be an individual's most current, most valid address. The following criteria may be used to select preferred address record:
    - Has the latest ADDRESS_PERIOD_START date
    - ADDRESS_PERIOD_END date is null (or, latest date if none are null)
    - ADDRESS_USE is HO (home)
    - ADDRESS_TYPE is PH (physical) or BO (both physical and postal)
- Implementers are permitted to include addresses that are from non-residential facilities such as long-term care centers, when available.

Table 6: Descriptions for PRIVATE_ADDRESS_HISTORY Attributes

Attribute	Description	Example
PATID	Arbitrary person-level identifier. Used to link across tables.	GREP9892
ADDRESSID	Arbitrary identifier for each unique address record.	ADD_GREP9892_1
ADDRESS_STREET	Primary address line (e.g., street name and number).	456 RIDGE RD.
ADDRESS_DETAIL	Remaining address details (e.g., suite, post office box, other details).	
ADDRESS_CITY	The name of the city, town, village or other community.	MILWAUKEE
ADDRESS_ZIP5	Five-digit postal code for the address.	34511
ADDRESS_STATE	State, as represented by 2-digit postal abbreviation.	WI
ADDRESS_TYPE	Type of address. Details of categorical definitions: • Postal: mailing address—PO boxes and care of addresses. • Physical: A physical address that can be visited. • Both: An address that is both physical and postal.	BO
ADDRESS_PREFERRED	Indicates whether this address is the preferred one for a given patient, address use, and address type within a given address period.	Y
ADDRESS_PERIOD_END	Date when address was no longer in use.	
ADDRESS_PERIOD_START	Initial date when the address is known to be in use.	11/13/2001
ADDRESS_USE	Purpose of the address. Details of categorical definitions: • Home: A communication address at a home. • Work: An office address. First choice for business-related contacts during business hours. • Temp: A temporary address • Old/Incorrect: This address is no longer in use (or was never correct but retained for records).	
ADDRESS_ZIP9	Nine-digit postal code for the address. (Do not include hyphens.)	345117889
RAW_ADDRESS_TEXT	Text representation of the address.	456 RIDGE RD.

4.3 QUESTION

The QUESTION[13] table is a reference table that defines the questions from questionnaires, screening tools, or assessments for which responses are present in the QUESTION_RESPONSE table. The QUESTION table is defined based on the FHIR Questionnaire resource[14], but limited to fields necessary for interpreting responses, not for administering the questionnaire. Responses to questions in the QUESTION table are captured in QUESTION_RESPONSE.

The purpose of the QUESTION table is to store relevant information about questions and the questionnaire that questions originate from. FHIR defines a questionnaire as an organized collection of questions intended to solicit information from patients, providers or other individuals involved in the healthcare domain. The FHIR Questionnaire resource includes the following questionnaire related information: the text in the questions to be asked, how the questions are ordered and grouped, any intervening instructional text, and what the constraints are on the allowed answers. The QUESTION table breaks the Questionnaire resource down into one row per question (Questionnaire.item) – fields related to the overall Questionnaire are duplicated and can be used for grouping. QUESTION includes a unique ID for each questionnaire (QUESTIONNAIRE_ID) and each question (QUESTION_ID) that can be referenced in QUESTION_RESPONSE.

Implementers should populate this table with questions from social needs screeners, topic-specific assessments (e.g., food insecurity assessment, activity of daily living assessment, fall risk assessment), and any other questionnaires that assess aspects of health. QUESTION can store item-level (i.e., single question) definitions as well as a total score associated with a set of questions (i.e., measure). Implementers should create one row for every question for which responses will be shared. For total scores, the implementer should add a row for each score, i.e., consider "total score" as an additional question. QUESTION captures text names of questionnaires and questions as well as standard codes (e.g., LOINC) for questionnaires and questions. Table 7 describes how implementers should populate the QUESTION table and provides examples.

Table 7: Descriptions for QUESTION Attributes

Attribute	Description	Example Value
QUESTION_ID	A primary key that uniquely identifies a row in the table.	12345
QUESTIONNAIRE_ID	An ID for the Questionnaire (group of questions) that this question is part of. For Questions that are not conceptually part of a group, implementers may choose to either use a unique QUESTIONNAIRE_ID per question, or a common QUESTIONNAIRE_ID shared by all questions that are not otherwise part of a group.	PHQ9

[13] The QUESTION and QUESTION_RESPONSE tables replace the PRO_CM table that has been deprecated. For more information about PRO_CM, refer to the V1.1 CODI Data Model Implementation Guide.

[14] FHIR Questionnaire Resource Website: https://build.fhir.org/questionnaire.html

Attribute	Description	Example Value
QUESTIONNAIRE_NAME	Name for this Questionnaire (human-friendly).	PHQ-9 quick depression assessment panel
QUESTIONNAIRE_CODE	Concept that represents the overall questionnaire.	44249-1
QUESTIONNAIRE_CODE_SYS	Vocabulary or standard present in QUESTIONNAIRE_CODE (e.g., LOINC, SNOMED, HCPCS).	LC
ITEM_ID	Unique id for this question within the questionnaire.	PHQ-9.1
CODE	Corresponding concept for this question in a terminology.	44250-9
CODE_SYS	Vocabulary or standard present in CODE (e.g., LOINC, SNOMED, HCPCS).	LC
TEXT	Primary text for the question.	Little interest or pleasure in doing things
ANSWER_OPTIONS	Human-readable list of possible answer options.	Not at all, Several days, More than half the days, Nearly every day
ENABLE_WHEN	Human-readable list of constraints that determine when this question is enabled. Suggested format: questionId operator value,… Example: 1= true,2!= false, etc.	PHQ-9.12 = false
REQUIRED	Whether the item must be included in data results.	true

4.4 QUESTION_RESPONSE

The QUESTION_RESPONSE[15] table is designed to store data reported by patients from questions administered verbally, or performed by the individual electronically or on paper that originate from a questionnaire. FHIR defines a questionnaire as an organized collection of questions intended to solicit information from patients, providers or other individuals involved in the healthcare domain. Other terms used to describe a questionnaire are survey, screener, and assessment.

Each row in the QUESTION_RESPONSE table is one response and also references the QUESTION table for the definition of the questions. The QUESTION_RESPONSE table is based on the FHIR QuestionnaireResponse resource,[16] broken down into one row per question (QuestionnaireResponse.item) – fields related to the overall QuestionnaireResponse are

[15] The QUESTION and QUESTION_RESPONSE table replace the PRO_CM table. For more information about PRO_CM, refer to the V1.1 CODI Data Model Implementation Guide.
[16] FHIR QuestionnaireResponse Resource: https://www.hl7.org/fhir/questionnaireresponse.html

duplicated and can be used for grouping. Table 8 describes how implementers should populate the QUESTION_RESPONSE table and provides examples.

Table 8: Descriptions for QUESTION_RESPONSE Attributes

Attribute	Description	Example Value
QUESTION_RESPONSE_ID	A primary key that uniquely identifies a row in the table.	6789
QUESTIONNAIRE_ID	Link back to the Questionnaire ID in the QUESTION table.	PHQ9
ITEM_ID	Link back to the Item ID in the QUESTION table.	PHQ-9.1
PATID	Arbitrary person-level identifier. Used to link across tables.	GREP9892
ENCOUNTERID	Arbitrary encounter-level identifier used to link across tables.	
AUTHORED	Date the answers were gathered.	2024-11-26T12:13
ANSWER_VALUE_STRING	A single response to the question, in string format.	Not at all
ANSWER_VALUE_CODE	The concept for this answer as represented in a terminology.	LA6568-5
ANSWER_VALUE_CODE_SYS	Vocabulary or standard present in ANSWER_VALUE_CODE (e.g., LOINC, SNOMED, HCPCS).	LC

4.5 ASSET_ENROLLMENT

The ASSET_ENROLLMENT table contains one record for each contiguous time period during which an individual is enrolled to receive assets. An asset is a resource transferred by an organization to an individual (i.e., meals, vouchers, or rides). The ASSET_ENROLLMENT table is designed to capture the assets that are *planned* for delivery whereas the ASSET_DELIVERY table captures the assets that were delivered. Differentiating between the enrollment for assets and an asset delivery event using distinct tables is appropriate because enrollment and delivery information may not be identical, and some implementers may have only enrollment records because delivery information is unavailable or not documented.

Implementers should populate the ASSET_ENROLLMENT table with individual-level enrollment information for a planned series of services that will be regularly delivered such as meal delivery four days per week, occurring for nine weeks. Because ASSET_ENROLLMENT records may persist over extended time periods, implementers will be required to update ASSET_ENROLLMENT records when enrollments end with an ASSET_END_DATE and DISPOSITION_REASON.

Table 9 provides implementation guidance for the ASSET_ENROLLMENT table.

Example Scenario 1

Imagine the following asset enrollment scenario to understand how implementers should populate ASSET_ENROLLMENT records. Suppose an individual was enrolled in a Home-delivered Nutrition Service, provided by Meals on Wheels of Central Maryland, that delivers two frozen meals a day, four days a week, starting in January for a year-long service. In this scenario:

- ASSET_NAME is “Meal Delivery Provider of Central Maryland”
- ASSET_DESCRIPTION is optional but may include a more detailed description of the asset.
- ASSET_TYPE is “Home-delivered Nutrition Service”
- ASSET_SUBTYPE is the string “frozen meals”
- ASSET_QUANTITY is the number “2”
- ASSET_UNIT contains the code for Meal “M”
- DELIVERY_FREQ is the number “4”
- DELIVERY_FREQ_UNIT is the code for Weekly “W”
- DELIVERY_START_DATE is “01/01/2024”
- DELIVERY_END_DATE is “12/31/2024”

Delivery Frequency

An implementer should choose the code that best represents the planned delivery frequency. For deliveries occurring less often than the selected frequency (e.g., daily, weekly, monthly), this number will be less than 1. For instance, there is a distinction between a monthly delivery versus a delivery every 28 days and a delivery four times a month versus once a week. If the planned delivery is:

- once per calendar month
    - DELIVERY_FREQUENCY_UNIT should contain the code for Monthly “M”
    - DELIVERY_FREQ is the number “1”
- every 28 days or every 4 weeks
    - DELIVERY_FREQ_UNIT should contain the code for “Weekly” “W”
    - DELIVERY_FREQ is the number “0.25” (i.e., once every four weeks)
- four times per calendar month
    - DELIVERY_FREQ_UNIT should contain the code for Monthly “M”
    - DELIVERY_FREQ is the number “4”
- once per week
    - DELIVERY_FREQUENCY_UNIT should contain the code for Weekly “W”
    - DELIVERY_FREQ is the number “1”

Multiple Assets

The ASSET_ENROLLMENT record is meant for just one asset type, subtype, and asset unit type. In scenarios where there are multiple kinds of assets delivered as part of the same program or at the same time, implementers should create multiple enrollment records; one enrollment record for each kind of asset. For example, if an individual is enrolled in a service that provides both frozen meals and fresh meals and transportation to a congregate meal venue

(e.g., rides to a community kitchen) then that information is described using one ASSET_ENROLLMENT record for the fresh meal delivery, one enrollment record for the frozen meal delivery, one enrollment record for the transportation service, and one enrollment record for the congregate meal. Implementers are encouraged to disaggregate asset enrollment types to represent only one type of service per record.

Asset Delivery Suspension

Service providers may interrupt enrollments and deliveries with "hold" periods where services are suspended for individuals, for various reasons, like hospitalization. Recognizing that, implementers may choose to terminate an enrollment and begin a new enrollment once asset delivery resumes.

Table 9: Descriptions for ASSET_ENROLLMENT Attributes

Attribute	Description	Example Values
ASSET_ENROLLMENT_ID	A primary key that uniquely identifies a row in the table.	05556
PATID	Arbitrary person-level identifier. Used to link across tables.	PAT659822
PROGRAMID	An optional link back to the program table to represent asset enrollments associated with a program.	
ASSET_NAME	A name of an asset.	Home Delivered Meals
ASSET_DESCRIPTION	A description of the asset.	The home delivered meals program provides fresh meals delivered to a client's home four days per week.
ASSET_DOMAIN	The Gravity Project SDOH domain that an asset is intended to address (e.g., food insecurity).	FD
ASSET_TYPE	Defines type of asset planned to be delivered (e.g., food, transportation).	HD
ASSET_SUBTYPE	Define subtype of asset planned to be delivered (e.g., hot meals, cold meals, frozen meals, shelf stable).	FROZEN
ENROLLMENT_DATE	A date on which the enrollment was performed.	01/01/2024
ASSET_QUANTITY	The planned number of asset units for each delivery.	2
ASSET_UNIT	The planned portion or specific type of asset or service for each delivery (e.g., meal).	M
DELIVERY_FREQ_UNIT	A unit of time used to describe how often an asset is delivered. For example, an asset delivered twice a week has a frequency of 2 and a unit of Weekly. An asset delivered every other week has a frequency of 0.5 and a unit of Weekly.	W

Attribute	Description	Example Values
DELIVERY_FREQ	The number of times an asset is delivered each unit of time.	4
DELIVERY_START_DATE	A date on which the individual who enrolled will start receiving asset.	03/01/2024
DELIVERY_END_DATE	A date on which the individual who enrolled will stop receiving the asset or service. This field may be blank if there is no planned end to the service.	
DISPOSITION_REASON	Reason program ended.	
PAYER_SOURCE	Asset payer source (e.g., waiver, grant).	GR
ORGANIZATIONID	A link back to the non-clinical organization that this enrollment belongs to.	MAC

4.6 ASSET_DELIVERY

The ASSET_DELIVERY table contains one record for every delivery or transfer of an asset from an organization to an individual. Implementers should populate the ASSET_DELIVERY with services delivered by their organization. Table 10 defines the attributes in the ASSET_DELIVERY table.

While the ASSET_ENROLLMENT table describes planned asset deliveries, the ASSET_DELIVERY table contains records for actual asset delivery events carried out by the service provider. Asset delivery events correspond to the planned deliveries in either an individual's asset enrollments or because of the program benefits associated with an individual's program enrollment. The asset delivery record will therefore almost always include a link to either an asset enrollment record (ASSET_ENROLLMENT_ID) or a program (PROGRAM ID). It is possible to populate both ASSET_ENROLLMENT_ID and PROGRAM_ID if appropriate and available. Implementers are discouraged from leaving both attributes blank.

Implementers should describe delivered assets in the ASSET_DELIVERY table using the same guidance as provided for ASSET_ENROLLMENT. Select an ASSET_TYPE from the allowed values, enter text for the ASSET_SUBTYPE, and select an ASSET_UNIT code from allowed asset unit values. Implementers are encouraged to match values for ASSET_TYPE, ASSET_SUBTYPE, and ASSET_UNIT that they used for the corresponding ASSET_ENROLLMENT records. Records in ASSET_DELIVERY may reflect one asset delivered on one day, or multiples of the same asset unit (e.g., meal) from the same service (e.g., Home Delivered Nutrition Service) delivered over a series of days.

Multiple Asset Deliveries

Individuals may be in receipt of multiple assets at the same time and as such have multiple ASSET_DELIVERY records for the same day. An individual may be receiving assets of different types from the same or different organizations such as meals from organization A and transportation vouchers from organization B. An individual may also be receiving multiple different assets of the same type (e.g., food) such as home delivered meals and groceries. Because ASSET_DELIVERY disallows the selection of multiple ASSET_TYPE values or ASSET_UNIT values, implementers should avoid combining multiple assets. For example, if an individual received a ride to a doctor's appointment and received a delivered meal, implementers should record the delivery of two assets: a meal and a ride, rather than one record for "meal and ride."

Table 10:Descriptions for ASSET_DELIVERY Attributes

Attribute	Description	Example Values
ASSET_DELIVERY_ID	A primary key that uniquely identifies a row in the table.	BH23391NDI
PATID	A link back to the demographic table.	PAT12345
PROGRAM ID	If applicable, the program for which this asset delivery is a part of.	
ASSET_ENROLLMENT_ID	If applicable, a link back to the asset enrollment that planned this asset delivery. An asset enrollment ID is optional but encouraged.	P45722NNRIF
ASSET_TYPE	Defines type of asset delivered (e.g., food, transportation).	HD
ASSET_SUBTYPE	Define subtype of asset delivered. For food insecurity, recommended values include fresh meals, frozen meals, and shelf stable.	“frozen meal”
SDOH_INTERVENTION_PROC	Code for social determinants of health procedures related to the provision of assets (Gravity).	464421000124107
SDOH_INTERVENTION_PROC_CODE_SYS	Vocabulary or standard used in the SDOH_INTERVENTION_PROC attribute. (e.g., LOINC, SNOMED); constrained to 5 code systems used by Gravity.	SM based on CODE_SYS valueset
DELIVERY_ORGANIZATIONID	A link back to the organization that delivered assets to the person. This organization may or may not be the same as the organization that enrolled the person to receive assets.	MAC
ASSET_QUANTITY	The total number of asset units delivered (e.g., total number of meals).	1
ASSET_UNIT	The portion or specific type of asset or service for each delivery (e.g., meal).	M
DELIVERY_START_DATE	The date or the first date of the date range the asset(s) was delivered.	05/13/2024
DELIVERY_END_DATE	The end date the asset(s) was delivered. For a single day, the entry may populate the same date for start and end.	05/13/2024

4.7 PROGRAM

The PROGRAM table contains one record for each program. This table will likely need to be manually implemented; that is, the required program information may not be stored in an information system. Since information about a program is not likely to change, in most circumstances, this table will need to be initially populated with programs offered by an organization and updated if new program offerings are added. Recognizing that organizations

may operate many programs unrelated to social needs and services, implementers should populate this table with one record per program for which PROGRAM_ENROLLMENTs will be populated. Table 11 describes how to implement the PROGRAM table attributes. Each location at which a program is administered, and each organization's version of a program constitutes a distinct program record. For example, each organization that administers a congregate meal or care transition program should appear separately in the PROGRAM table.

The PROGRAM table captures an implementer's best understanding of how a program is administered and for what purpose. The attributes with the PROGRAM_ and AIM_ prefixes apply to every program. The AFFILIATED_PROGRAM attribute provides a way to document that a given program is affiliated with an encompassing program or program category.

Duration and Frequency

The attributes with the PRESCRIBED_ prefix only apply to those programs with a predefined frequency of interaction and duration, such as a program that lasts for 10 weeks and meets twice a week for two hours each time. These attributes allow implementers to describe the programs intended dose and intensity; these attributes should be blank for programs without a predefined dose.

When completing PRESCRIBED_SESSION_FREQUENCY, implementers should calculate the number of sessions that are intended to occur in a week. For sessions occurring less often than weekly, this number will be less than 1. For instance, if the prescribed session frequency is:

- daily (Monday-Thursday)
    - PRESCRIBED_SESSION_FREQUENCY is the number "4" for 4 sessions per week.
- every 28 days or every four weeks
    - PRESCRIBED_SESSION_FREQUENCY is the number "0.25" for 0.25 sessions every week (i.e., once every four weeks).

Location

The attributes with the LOCATION_ prefix describe the location at which the program is typically administered. Three variants of location are supported: address, geospatial coordinates (latitude and longitude), and geocode (typically census tract). Implementers should provide all three location attribute variants for each program, if possible. Some programs are designed to be attended online or at home with program materials. The LOCATION_ADDRESS for these programs should be indicated with the string "Virtual" and latitude, longitude, and geocode may be null.

Systematically Missing Session Information

The attributes with the SESSION_OMISSION_ prefix describe the circumstances under which session information is missing for the program. For example, some clinical programs only record sessions with a clinical component. The sessions lacking a clinical component are not documented and therefore do not appear in the SESSION table. Other programs exhibit less systematic omissions (e.g., because attendance is sometimes captured on paper). These attributes are included to help researchers better decide how to handle missing session information.

Table 11: Descriptions for PROGRAM Attributes

Attribute	Description	Example
PROGRAMID	A unique identifier that identifies a program and is a primary key that uniquely identifies a row in the table.	CTC
PROGRAM_NAME	A name of the program.	Care Transition Program
PROGRAM_DESCRIPTION	A description of the program.	Post-hospitalization outreach, screening and navigation
PROGRAM_SETTING	A setting in which the program is offered (clinical or community).	
PROGRAM_ORGANIZATIONID	A link back to the organization that is offering the program.	MAC
AFFILIATED_PROGRAMID	A parent program that encompasses this and other programs that are often prescribed together. A parent program may not have any sessions of its own, nor a physical location but may serve only to encompass affiliated programs.	
AIM_NUTRITION	True if the aim of the program includes improving nutrition.	
AIM_ACTIVITY	True if the aim of the program includes improving physical activity.	
AIM_WEIGHT	True if the aim of the program includes improving weight status.	
AIM_SDOH	True if the aim of the program is to address social determinants of health.	Y
PRESCRIBED_TOTAL_DOSE	A total amount of time (in hours) an individual should spend in the program. This field should equal DURATION x FREQUENCY x LENGTH (6 *Weeks* times 1 *Session per week* times 1 *Hour per session*).	6
PRESCRIBED_PROGRAM_DURATION	A measure of the time (in weeks) from start to finish.	6
PRESCRIBED_SESSION_FREQUENCY	A number of sessions delivered each week.	1

Attribute	Description	Example
PRESCRIBED_SESSION_LENGTH	A number of hours delivered each session.	1
LOCATION_ADDRESS	A primary location at which this program's sessions are administered, expressed as an address. If the program is designed for participants to participate entirely online or at home, then the location value should be the string "Virtual."	Virtual
LOCATION_LATITUDE	A latitude of the corresponding address location.	
LOCATION_LONGITUDE	A latitude of the corresponding address location.	
LOCATION_GEOCODE	A primary location at which this program's sessions are administered, expressed as a geocode.	
LOCATION_BOUNDARY_YEAR	A census year for which the corresponding geocode location applies.	
LOCATION_GEOLEVEL	A specificity of the geocode location. This can be assessed using logic that considers the length of the GEOCODE value (2 characters for state; 5 characters for county; 11 characters for census tract).	
SESSION_OMISSION_PERCENT	A numeric estimate of the percentage of all sessions missing from the SESSION table (based on intended dose) for this program; 0% indicates a belief that the session information is fully populated.	
SESSION_OMISSION_DESCRIPTION	A description of the circumstances under which session information for this program is missing; this field is required when the omission percent is greater than 0%.	
SESSION_OMISSION_SYSTEMATIC	True if session information for this program is systematically missing (e.g., because only half of the sessions are documented).	

4.8 PROGRAM_ENROLLMENT

The PROGRAM_ENROLLMENT table captures information about an individual's enrollment in a program (described in Section 4.7). Enrolling in a program does not mean the individual participated in the program, only that they were registered to participate. The individual's participation is captured in the SESSION table (see Section 4.9) if attendance or encounter information is available. Implementers should populate the PROGRAM_ENROLLMENT table with individual-level enrollment records for programs listed in the PROGRAM table. The PROGRAM_ENROLLMENT table also captures the individual's completion of the program. How a program defines completion is left to the discretion of each program. Researchers should consult the program's documentation to understand how best to interpret program completion. If an individual disenrolls from the program, implementers should populate the disposition description to document the circumstances involved. Table 12 provides implementation guidance for the PROGRAM_ENROLLMENT table.

Table 12: Descriptions for PROGRAM_ENROLLMENT Attributes

Attribute	Description	Example
PROGRAM_ENROLLMENT_ID	A primary key that uniquely identifies a row in the table.	PL33472MC
PATID	Arbitrary person-level identifier. Used to link across tables.	MRAG8308
PROGRAMID	A link back to the program this enrollment belongs to.	CTC
ORGANIZATIONID	A link back to the non-clinical organization that this enrollment belongs to.	MAC
REFERRALID	A link back to the referral that this enrollment belongs to.	
ENROLLMENT_DATE	A date on which the enrollment was performed.	01/03/2024
COMPLETION_DATE	A date on which the individual who enrolled completed the program.	05/03/2024
DISPOSITION_REASON	Reason program ended.	ND
PAYER_SOURCE	Program payer source (e.g., waiver, grant).	GR

4.9 SESSION

The SESSION table contains one record for each interaction between an individual and an organization or program representative as part of their enrollment in a program. Implementers should populate the SESSION table with records for individuals' attendance or participation in programs.

In the SESSION table, there is one record for each time an individual attends the program. For example, an individual who completes a program that meets weekly for 10 weeks should have 10 distinct SESSION records. Sessions could be classes, congregate meals, or group or 1:1 engagements with peers or professionals. Sessions may include one or multiple activities and

may be structured with a defined curriculum or unstructured. The SESSION table includes several process-related attributes (SCREENING, COUNSELING, and those with the INTERVENTION_ prefixes). In some cases, the values of these attributes need to be established based on local program knowledge as opposed to what is present in an organization's information system.

The DOSE attribute indicates the amount of time spent interacting with the individual (in hours). This attribute should only be populated based on what is documented in an information system. If the duration of the session is not documented, the DOSE attribute should remain empty. Table 13 describes the attributes in the SESSION table.

Table 13: Descriptions for SESSION Attributes

Attribute	Description	Example
SESSIONID	A primary key that uniquely identifies a row in the table.	PRLEKSH33
PATID	Arbitrary person-level identifier. Used to link across tables.	MRAG8308
ENCOUNTERID	A link back to the encounter this session corresponds to (if any).	
PROVIDERID	A provider primarily responsible for this session.	
PROGRAMID	A link back to the program this session belongs to (if any).	CTC
PROGRAM_ENROLLMENT_ID	A link back to the enrollment event this session belongs to.	
SESSION_DATE	A date on which the session was conducted.	03/03/24
SESSION_MODE	An indication of the way the session was delivered (e.g., individual, group, phone).	PH
SCREENING	True if the session included any assessment of health-related factors such as social needs, physical activity, nutrition, or sleep.	Y
COUNSELING	True if the session included any advice or direction regarding lifestyle or behaviors such as physical activity, nutrition, or sleep.	Y
INTERVENTION_ACTIVITY	True if the session included performing at least moderate physical activity; moderate activity requires a moderate amount of effort (5-6 on a scale of 0 to 10) and noticeably accelerates the heart rate and breathing.	NI

Attribute	Description	Example
INTERVENTION_NUTRITION	True if the session included an activity designed to improve nutrition.	NI
INTERVENTION_NAVIGATION	True if the session included a navigational service to access benefits or to overcome barriers to care.	Y
INTERVENTION_SDOH	True if the sessions associated with this curriculum include an activity designed to address a social determinant of health	Y
DOSE	A measure of the amount of time spent on this encounter. Researchers can compare the total dose to the prescribed total dose to assess the extent to which an individual completed a program.	
CURRICULUM_COMPONENT_ID	A link back to the curriculum component (if any) associated with this session.	CTC_30DAY

4.10 CURRICULUM_COMPONENT

The CURRICULUM_COMPONENT table is a reference table that defines the standard elements of a program in order to describe what is intended to happen throughout the course of the program. The SESSION table describes what has been documented as having transpired. The CURRICULUM_COMPONENT table provides insight into what likely happened when session information is missing or incomplete. The SESSION table includes an optional reference to CURRICULUM_COMPONENT; however, sessions are not required to have CURRICULUM_COMPONENTs. This table supports both a fixed curriculum, in which the components are ordered using SESSION_INDEX, and a recurring curriculum, in which the components repeat. Repeating components are documented with a combination of SESSION_FREQ and SESSION_FREQ_UNIT, as described in Section 4.7 above. Table 14 describes how to implement attributes in the CURRICULUM_COMPONENT table.

Programs with a Standard Curriculum

For any programs that have standard elements, implementers may describe the activities included in that CURRICULUM_COMPONENT record by indicating if that CURRICULUM_COMPONENT included screening, counseling, physical activity, nutrition, navigation, or addressed social needs. One CURRICULUM_COMPONENT record can and often includes multiple activities. Because a record in the SESSION table may reference only one CURRICULUM_COMPONENT record, implementers are encouraged to consolidate CURRICULUM_COMPONENT records to encompass the totality of activities that are accomplished in the session rather than creating distinct records for each activity conducted during a session. When standard, implementers may also record the CURRICULUM_COMPONENT’s duration in the DOSE field.

Program examples with standard CURRICULUM_COMPONENTs are a care transition or chronic disease self-management program. For a care transitions program, the standard 30, 60, 90 and 120 day follow up telephone calls would be captured as four CURRICULUM_COMPONENT records. For a 6-week chronic disease self-management program, each week's curriculum would be documented as a CURRICULUM_COMPONENT record for a total of 6 records.

Table 14: Descriptions for CURRICULUM_COMPONENT Attributes

Attribute	Description	Example
CURRICULUM_COMPONENT_ID	A primary key that uniquely identifies a row in the table.	CTC30DAY
PROGRAMID	A link back to the program this component of a curriculum belongs to.	CTC
SESSION_INDEX	An ordinal used to establish a total ordering on the sessions within a fixed curriculum.	1
SESSION_FREQ	A number of times a session is administered each unit of time.	1
SESSION_FREQ_UNIT	A unit of time used to describe how often a session is administered. For example, a session administered twice a week has a frequency of 2 and a unit of Weekly. A session administered every other week has a frequency of 0.5 and a unit of Weekly.	O
SCREENING	True if the session included any assessment of health-related factors such as social needs, physical activity, nutrition, or sleep.	Y
COUNSELING	True if the session included any advice or direction regarding lifestyle or behaviors such as physical activity, nutrition, or sleep.	NI
INTERVENTION_ACTIVITY	True if the sessions associated with this curriculum include performing at least moderate physical activity; moderate activity requires a moderate amount of effort (5-6 on a scale of 0 to 10) and noticeably accelerates the heart rate and breathing.	N
INTERVENTION_NUTRITION	True if the sessions associated with this curriculum include an activity designed to improve nutrition.	NI
INTERVENTION_NAVIGATION	True if the sessions associated with this curriculum include a navigational service to access benefits or to overcome barriers to care.	Y
INTERVENTION_SDOH	True if the sessions associated with this curriculum include an activity designed to address a social determinant of health.	Y
DOSE	A measure of the amount of time sessions associated with this curriculum are expected to last.	

4.11 ORGANIZATION

The ORGANIZATION table is a reference table and contains one record for each organization referenced in the ASSET_ENROLLMENT, ASSET_DELIVERY, PROGRAM, PROGRAM_ENROLLMENT, or REFERRAL table. Implementers should create a record in ORGANIZATION for any organization collecting question responses, sending or receiving referrals, offering programs, and delivering assets. This table describes organizations by contact information. Implementers should not include organizations in the ORGANIZATION table who do not have records in any other tables in the SDE. It is likely for many implementers that the ORGANIZATION table will include only one row for their organization. Table 15 describes the attributes in the ORGANIZATION table.

Table 15: Descriptions for ORGANIZATION Attributes

Attribute	Description	Example
ORGANIZATIONID	A primary key that uniquely identifies a row in the table.	MOW
ORGANIZATION_NAME	Name of organization.	Meals on Wheels
ORGANIZATION_ADDRESS	Address of organization.	123 Bay rd.
ORGANIZATION_CITY	City of organization.	Sheboygan
ORGANZIATION_STATE	State, as represented by 2-digit postal abbreviation.	Wi
ORGANIZATION_ZIP	5-digit postal code for the address.	31445
ORGANIZATION_PHONE	Phone number of organization.	999-222-3333
ORGANIZATION_TYPE	For clinical organizations, use the CMS Certification Number (CCN); each implementing network will need to choose a representative CCN for its clinical data partners. For community organizations, each implementing network will need to establish a set of community organization codes. These additional codes should include at least one letter so that they do not conflict with CCNs.	AAA1

4.12 REFERRAL

The REFERRAL table contains one record for each outgoing or incoming electronic referral for programs, social services, or clinical services. Implementers should populate this table with individual-level referrals generated by or received by their information systems for social needs screening or for social services and programs. Implementers should not engineer REFERRAL records for referrals received through alternate pathways (e.g., telephone, fax, or email) or self-referrals. Table 16 provides implementation guidance for the REFERRAL table.

There are two key concepts that the REFERRAL table seeks to capture: who the referral was between and what the referral was for. REFERRAL contains multiple attributes to capture the source and destination organization that describe who the referral was between. These fields may also be useful to link outgoing referrals with incoming referrals to evaluate whether a referral successfully connected an individual with a program or service – sometimes referred to

as closing the loop. REFERRAL accepts outgoing referrals without a corresponding incoming referral and vice versa. If a referral is internal to an organization (incoming and outgoing within the same organization), those referrals may be included in the REFERRAL table as two records: one outgoing referral and a second incoming referral and may reference the same organization in source and destination fields. The DIRECTION attribute indicates if the record represents the implementer initiating a referral (outgoing) or receiving a referral (incoming). DESTINATION_SPECIALTY, DESTINATION_ASSET_TYPE_CODE, AND DESTINATION_PROGRAM_ID are the attributes that may be used to describe what the referral is for.

Clinical Referrals

For clinical organizations sending and receiving referrals, SOURCE_ORGANIZATIONID and DESTINATION_ORGANIZATIONID identify organizations sending and receiving referrals using CMS Certification Numbers. For these, implementers will need to ensure that sending and receiving organizations are accurately represented in ORGANIZATION. Similarly, SOURCE_PROVIDERID is intended to identify specific healthcare providers within a healthcare organization. DESTINATION_SPECIALTY should be populated for referrals for specialty clinical services.

Non-Clinical Referrals

For referrals sent from or to non-clinical entities, DESTINATION_SPECIALTY and SOURCE_PROVIDERID may be disregarded and implementers may use SOURCE_ORGANIZATIONID and DESTINATION_ORGANIZATIONID to identify sending and receiving organizations.

DESTINATION_ASSET_TYPE_CODE should be populated for referrals for specific social services and allows implementers to select the appropriate value from a Gravity value set. DESTINATION_PROGRAMID should be populated for referrals to programs like falls prevention or chronic disease self-management.

Table 16: Descriptions for REFERRAL Attributes

Attribute	Description	Example
REFERRALID	A primary key that uniquely identifies a row in the table.	ABC1234
PATID	Arbitrary person-level identifier. Used to link across tables.	KRAMA33
ENCOUNTERID	A link back to the encounter table, if the referral can be unambiguously associated with an encounter.	
DIRECTION	An indication of whether the referral was incoming or outgoing.	INCOMING
REFERRAL_DATE	A date the referral was made.	03/03/24
REFERRAL_STATUS	A final disposition of the referral.	
REFERRAL_PRIOR_AUTH	An indication of whether prior authorization was required for the referral.	
SOURCE_PROVIDERID	A provider responsible for initiating this referral.	

Attribute	Description	Example
SOURCE_ORGANIZATIONID	A link back to the non-clinical organization that initiated the referral (if applicable).	
DESTINATION_ORGANIZATIONID	A link back to the non-clinical organization to which the referral was sent (if applicable).	MAC
DESTINATION_SPECIALTY	A clinical specialty for which the patient is being referred.	
DESTINATION_ASSET_TYPE_CODE	Defines type of asset requested in referral (e.g., meals, rides, vouchers).	
DESTINATION_ASSET_CODE_SYS	Vocabulary or standard present in DESTINATION_ASSET_TYPE_CODE.	
DESTINATION_PROGRAMID	A program to which an individual is being referred.	CTC

5 Generating and Sharing the Structured Data Extract

Implementers will apply the SDE table and attribute structure through an ETL process that results in a SDE file that can be shared with an external organization. This section provides guidance for generating and sharing the SDE.

5.1 Generating a Structured Data Extract

Generating an SDE from an information system involves several key steps to ensure the process is efficient and the data retrieved and included in the SDE is accurate and relevant.

Define Objective and Requirements

First, it is essential to clearly define the objectives and requirements of the data extract. Implementers should identify the specific information systems, data fields, tables, and records needed, as well as any filters, mappings, or conditions that should be applied to normalize the data to the data model.

Define Approach

Implementers should set an approach to sharing SDE that informs how the data are extracted. The following approaches are recommended:

- **Historical to incremental data extracts**: Implementers will first extract a full historical SDE that includes data extending back to the designated start date and then transition to sharing monthly increments of data that are smaller and include only data about activities that occurred since the last extract. This approach may require a modification of dates in the data extract process.
- **Expanding historical data extracts**: Implementers will extract a full historical SDE that includes data extending back to the designated start date. With subsequent extracts, the full historical data will be included and will grow slightly in size with the addition of newly created data.

Extract, Transform, and Load Data

Once the SDE requirements are established, the next step is to access data stored in the information system, which may involve logging into a database management system or using a specialized software interface. For information systems without backend access, implementers may plan to generate the SDE from manually or automatically generated extracts or reports (e.g., comma separated values [CSV] or text [TXT] file). When pulling from reports or extracts, the next step is to generate the reports or extract files needed for the SDE. Figure 3 displays how ETL can be accomplished to generate an SDE:

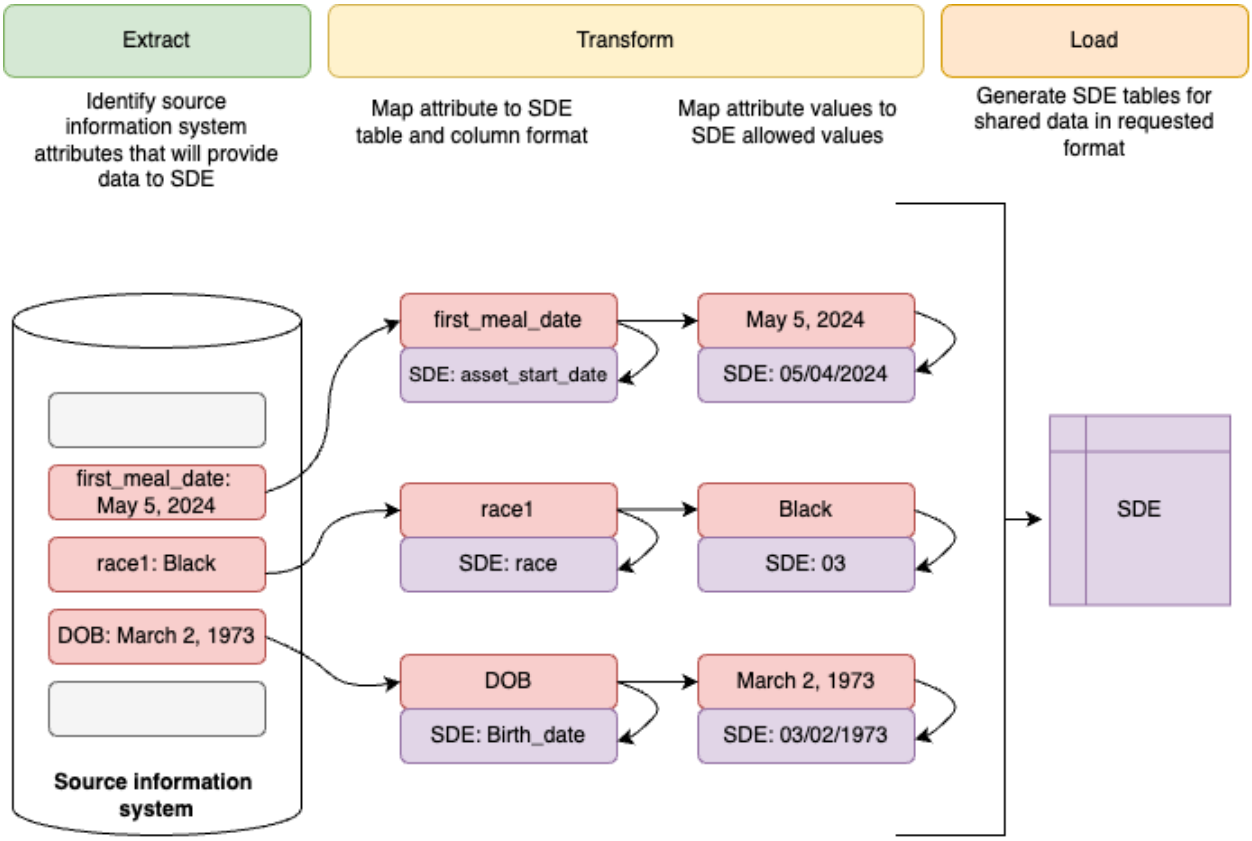

Figure 3: Extract, Transform, and Load Diagram

Depending on the system, users might employ Structured Query Language (SQL) to write queries that specify the data to be extracted. For instance, a simple SQL query might look like SELECT * FROM customers WHERE DELIVERY_START_DATE > '2022-01-01'; to retrieve all client records enrolled in after January 1, 2022. Implementers should use the predefined start date (see section 3.3).

After constructing the query, implementers should test the query or extract to ensure it returns the correct and complete data and that the transformations have been implemented correctly. This can be done by running the query on a small subset of the database or using a development environment that mirrors the production system. Implementers should check for any errors or warnings generated from the extraction process. Recommended overall checks include:

- Are all of the expected SDE tables present?
- Do SDE tables include records that have expected visible patterns of missingness?

- Do SDE tables include an expected volume of records?
- Do dates in the SDE tables align with the extraction data parameters?

Implementers are encouraged to define and implement additional checks to validate table and attribute specific elements based on their service delivery model and information systems.

Once the query is validated, the data extraction process can proceed in full, generating a complete historical SDE file. The extracted data can be exported in various formats; implementers are encouraged to share data in a CSV or JSON format.

Security and Privacy

It is also important to consider data security and privacy during extraction, ensuring that sensitive information is handled appropriately and that access controls are in place. Because the SDE includes PII, implementers should extract SDE files to a secure location accessible only by other authorized users.

ETL Process Documentation

Implementers are encouraged to document their extraction process to ensure that the process can be consistently replicated. This is particularly important in environments where data needs to be extracted regularly or by different team members. Clear documentation provides a step-by-step guide that helps maintain the integrity and accuracy of the SDE over time, reducing the risk of errors or discrepancies. It also serves as a valuable reference for troubleshooting and refining the process, enabling quick identification and resolution of any issues that may arise. Process documentation enhances transparency and accountability within an organization. By clearly outlining the methods and tools used for data extraction, as well as any assumptions or parameters applied, stakeholders can better understand the origins and context of the data. Additionally, well-documented processes facilitate knowledge transfer and onboarding, allowing new team members to quickly get up to speed and contribute effectively.

As part of extraction, implementers should implement a naming convention where the SDE files are named consistently to avoid confusion. A date should be included in the file name so that old files are not written over with subsequent extracts. Implementers are encouraged to put SDE file extracts in a folder that includes the date.

Over time, multiple SDEs will be generated and shared. Implementers are encouraged to create a log of SDE files that includes:

- **Timestamp**: Record the date and time when the data extraction was initiated and completed.
- **User Information**: Document the identity of the user or system account that performed the extraction.
- **Data Source**: Specify the database, table, or system from which the data was extracted.
- **Query or Extraction Criteria**: Define the criteria for data extraction.
- **Extracted Data Description**: Summarize the data extracted, including the number of records, the fields or columns included, and any filters or conditions applied.
- **Output Format**: Indicate the format in which the data was exported.
- **Destination**: Document where the extracted data was stored or sent.
- **Purpose**: Note the reason for the data extraction, whether it is for reporting, analysis, compliance, or another purpose.

- **Errors and Warnings**: Log any errors, warnings, or issues encountered during the extraction process. This information is crucial for troubleshooting and improving the extraction process over time.

Implementers should retain a copy of all SDE files, as feasible. Over time, the burden to store multiple large SDE files may lead implementers to archive or destroy old SDE files after an extended period of time. Following the recommended ETL steps, implementers may proceed to sharing their SDE data.

5.2 Sharing a Structured Data Extract

Sharing the SDE includes securely sending or transferring the extract file to an authorized recipient. Implementers are encouraged to use access control best practices, only sharing the extract file with individuals who need access to it and are authorized to receive it. Use role-based access controls if possible, and ideally, ensure that the recipient is authenticated before providing access to the data.

When sending a data extract to another party, it's crucial to ensure that the data is transmitted securely to protect sensitive information from unauthorized access. Use strong encryption methods to encrypt the data before sending it. Use secure communication channels such as HTTPS, SFTP (Secure File Transfer Protocol), or VPN (Virtual Private Network) to transmit the data. Implementers may protect the data extract file with a strong password and should send the password through a different communication channel than the data extract itself (e.g., if you send the data via encrypted email or SFTP, send the password via secure messaging app).

Implementers are expected to ensure that the data transfer complies with relevant regulations and standards. Data sharing agreements that outline the responsibilities and obligations of the sending and receiving parties regarding data security are recommended.

6 Additional Resources

6.1 Requesting Changes

SDE implementers are likely to find limitations with this Implementation Guide or the DD and can suggest changes. The process for requesting changes and approval will be documented in the CODI playbook.[17]

CODI Data Model or SDE changes will follow typical semantic versioning. Changes that are backwards compatible increment the minor version number of the Implementation Guide (e.g., from 1.4 to 1.5). Changes that are not backwards compatible (e.g., replacing an ancillary table with a table from CODI Data Model) will increment the major version number (e.g., from 3.3 to 4.0). Changes in the implementation guidance that do not require any data model change such as revising guidance to improve clarity, will increment the iteration number of the minor version (e.g., from 4.1.2 to 4.1.3).

[17] The CODI playbook will be available at https://mitre.github.io/CODI/ by August 2025.

Appendix A CDC CODI Data Models

The CDC CODI data models were developed as an extension of the PCORnet Common Data Model,[18] a standard data model representing anonymized patient-level data for research.

The CODI Research Data Model (RDM) represents data about health care services, healthcare utilization, and community-based program and service enrollment and participation. The CODI Recording Linkage Data Model (RLDM) is needed when using PPRL to match an individual's records across different data owners without sharing identifiers.

Figure A-1 illustrates the CODI data models. A unified implementation guide describing both data models in detail is available for reference.[19]

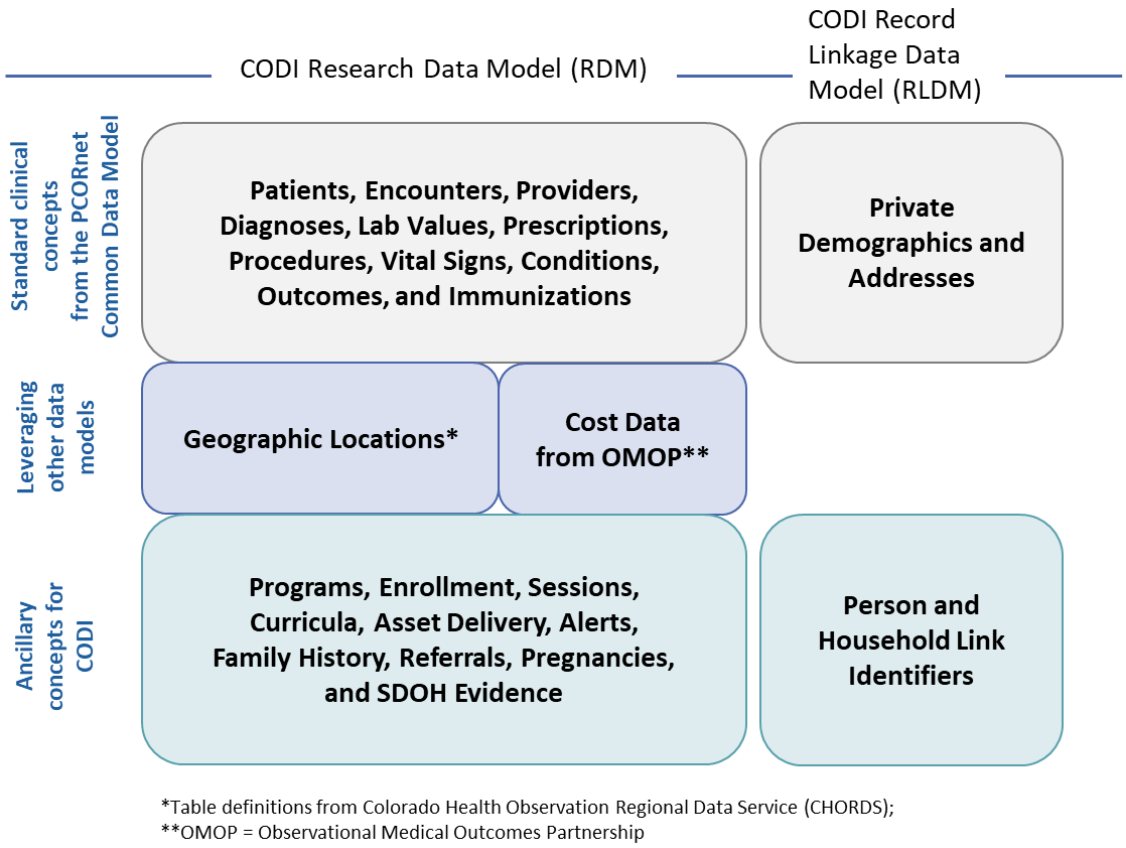

Figure A-1. Overview of CODI Research and Record Linkage Data Models

[18] PCORnet Common Data Model: https://pcornet.org/wp-content/uploads/2023/04/PCORnet-Common-Data-Model-v61-2023_04_031.pdf

[19] Version 4 (March 2023) is the latest version of the CODI Data Models Implementation Guide (DM IG).

Appendix B Additional Implementation Guidance

B.1 Relational Database

A relational database is a set of tables each with one or more rows and one or more columns. A row holds a data record (e.g., an office visit record), and a column holds values for one attribute for all the records (e.g., office visit dates). Each table cell holds one attribute value ("June 20, 2024") for one data record (e.g., Fred's office visit). Each table definition designates a column (i.e., attribute), often called "ID," as its unique record key. Data records are linked to one another by having the same cell values in their respective record keys. In this way, Fred's office visit record can be linked to Fred's vital signs, for example.

Implementers can export and import the contents of a relational database using a text file format called CSV. The first row of a CSV may contain comma separated column names that provide the attribute order of the comma separated values in the remaining rows. This implementation guide explains how an implementing organization should represent their exported data in a CSV file so that the receiving organization can properly import those data into its relational database.

B.2 Data Standards and Value Sets

Each organization stores data and information in a unique way and that makes combining the same type of data across organizations difficult. A data standard is a type of standard, which is an agreed upon approach to allow for consistent measurement, qualification, or exchange of an object, process, or unit of information. Data standards refer to methods of organizing, documenting, and formatting data in order to aid in data aggregation, sharing, and reuse. There are many data standards, and data standards can be generated by a research community (e.g., Observational Health Sciences and Informatics), a governmental organization (e.g., International Organization for Standardization), or other large organizations.[20] There are many data standards for health data and for a given type of health data, such as diagnostic codes, multiple standards are applicable.

B.3 Reference Tables

A reference table is a table that contains standardized, defined values that are used across other tables in the database. These values serve as a reference for data in other tables, promoting consistency, reducing redundancy, and enhancing data integrity.

The ORGANIZATION, PROGRAM, and CURRICULUM_COMPONENT tables are SDE reference tables that are unlikely to be populated directly from a source information system. Instead, these tables need to be populated by implementers during the ETL process.

Furthermore, implementers are encouraged to test the referential integrity of these tables. This means checking that the ORGANIZATION, PROGRAM, and CURRICULUM_COMPONENT tables correctly link to the tables that reference them, such as PROGRAM_ENROLLMENT and SESSION. This ensures the data across different tables is consistent and accurate.

[20] National Library of Medicine definition of data standard: https://www.nnlm.gov/guides/data-glossary/data-standards#:~:text=Definition,process%2C%20or%20unit%20of%20information.

B.4 Privacy Best Practices

Implementers are encouraged to implement the following best practices to protect the privacy of their patients and clients.

- **Limit Access**: Only authorized individuals should have access to PII and sensitive data. This can be achieved through access controls, such as passwords or physical controls like locked doors and secured physical spaces.
- **Data Minimization**: Collect and store only the PII that is necessary for the task at hand. Avoid collecting unnecessary information.
- **Training**: Provide regular training to all staff members who handle PII. They should understand their responsibilities under data protection laws and regulations.
- **Regular Audits**: Conduct regular audits to ensure ongoing compliance with data protection regulations and standards.
- **Data Breach Procedures**: Have procedures in place for managing data breaches, including identifying and reporting breaches, as well as measures to mitigate the impact of any breaches.
- **Secure Disposal**: When no longer needed, PII should be securely disposed of, such as through shredding, degaussing, or secure digital deletion.

B.5 Social Needs Categories and The Gravity Project

There are national efforts to standardize the way clinical organizations and CBOs assess social needs (i.e. SDOH or HRSN) among the populations they serve.[21] Interviews with CBOs and clinical partners revealed that data collection instruments to detect social needs and assess specific social needs vary across organization and are evolving. Therefore, SDOH data are collected and represented in many forms across organizations.

The Gravity Project[22] is an HL7 Fast Healthcare Interoperability Resources[23] accelerator project for harmonizing social risk factor data and improving electronic health information interoperability.

As part of Gravity's ongoing harmonization effort, a broad set of stakeholder groups have categorized data elements from many existing SDOH tools used for screening, diagnosis, goal setting, and interventions. They call these categories SDOH Domains.[24] For example, Gravity has categorized questions from the Hunger Vital Sign[25] screening tool to their domain "Food Insecurity" so that they can compare those questions and codes to the food insecurity questions and codes from PRAPARE,[26] another screening tool.

Gravity's Terminology Workstream Dashboard in the HL7 Confluence site has a spreadsheet for each of Gravity's SDOH domains, containing screening questions, clinical codes, and data elements that they have mapped to domains. This is work in progress and will continue to be updated after the major release dates of any standards that they have mapped (e.g., SNOMED: March and September, ICD: October, LOINC: August and February).

---

[21] Agency for Healthcare Research and Quality, https://www.ahrq.gov/sdoh/index.html
[22] The Gravity Project: https://www.hl7.org/gravity/
[23] What is FHIR: https://www.healthit.gov/sites/default/files/2019-08/ONCFHIRFSWhatIsFHIR.pdf
[24] Gravity SDOH Data Elements: https://confluence.hl7.org/display/GRAV/SDOH+Data+Elements+And+Status
[25] Hunger vital sign: https://childrenshealthwatch.org/public-policy/hunger-vital-sign/
[26] Protocol for Responding to and Assessing Patients' Assets, Risks, and Experiences (PRAPARE)

Appendix C Acronyms

Term	Definition
AAA	Area Agency on Aging
ACL	Administration for Community Living
ADL	Activity of Daily Living
CAT	Computer Adaptive Testing
CBO	Community-Based Organization
CCN	CMS Certification Number
CDC	Centers for Disease Control and Prevention
CMS	Centers for Medicare & Medicaid Services
CODI	Clinical and Community Data Initiative
CSV	Comma Separated Value
DD	Data Dictionary
ETL	Extract–Transform–Load
FFRDC	Federally Funded Research and Development Center
FHIR	Fast Healthcare Interoperability Resources
HIE	Health Information Exchange
HL7	Health Level Seven International
ICD	International Clinical Diagnosis
LOINC	Logical Observation Identifiers, Names, and Codes
PCORnet	Patient Centered Outcomes Research Network
PII	Personally Identifiable Information
PPRL	Privacy-Preserving Record Linkage
PRO	Patient-Reported Outcome
SDE	Structured Data Extract
SDOH	Social Determinants of Health
SNOMED	Systematized Nomenclature of Human Medicine
SQL	Structured Query Language
VSAC	Value Set Authority Center