Sentinel Common Data Model (SCDM) v8.0.0: Frequently Asked Questions

Project Purpose: The SCDM v8.0.0 Implementation project incorporates several major enhancements including: addition of a Prescribing table and international coding terminologies; modifications to Encounter, Diagnosis and Procedure tables that reflect changes to provider, facility and provider specialty variables; and metrics quantifying efficiency improvements gained through these enhancements.

Below is a list of Frequently Asked Questions to orient and guide users regarding SCDM v8.0.0. If you have additional questions about SCDM v8.0.0, please contact [info@sentinelsystem.org]. For general SCDM questions please see the "How Sentinel Gets Its Data" webpage.

1. **What were the goals of this model change?**
a. Increase data capture.
b. Improve organizational precision of data to facilitate medical chart validation.
c. Broaden model accessibility to international Data Partners.
d. Introduce run-time efficiencies for reading and processing SCDM data.

2. **How does SCDM v8.0.0 increase capture of data?**
a. A Prescribing table was added to include data on prescriptions given to patients. Prescribing data are distinct from outpatient drug dispensings (for more detail, please see Question 8 below).
b. New Rx_CodeType and Rx variables were added to the Dispensing table to replace National Drug Codes (NDC) and allow for additional drug code types. Further, additional values were added to the Dx_CodeType variable in the Diagnosis table, and to the Px_CodeType variable in the Procedure table, to render the model extensible to international Data Partners.
c. Provider and Facility tables were created and variables such as ProviderID, Specialty, and Facility_Location were added to enable capture of Provider and Facility details.
d. Reference tables were created to facilitate adding laboratory test types to the Laboratory Results; COVID-19 diagnostic laboratory results were the first test types added.

3. **How does SCDM v8.0.0 improve organizational precision of data?**
a. The Provider and Facility_Location variables were removed from the SCDM Encounter table and relocated to the newly created Provider and Facility tables, de-coupling Provider and Facility_Location constructs from individual Encounters.
b. Provider Medical Specialty was added to the new Provider table.
c. The Diagnosis and Procedure tables allow for the capture of multiple providers who record diagnoses or perform procedures within the same Encounter.
d. The Source variable within the Cause of Death table has enhanced labeling.
e. Improved clarity on what qualifies as an Encounter.
f. Within the Encounter table, guidance was added to signify an "Ongoing Inpatient Encounter:" the Discharge Date variable is to be populated with a special missing value (e.g., S within SAS 9.4).

4. What changes make SCDM v8.0.0 accessible to international Data Partners?
a. The Dispensing table has been updated to allow for code types beyond NDC (e.g., Systemized Nomenclature of Medicine (SNOMED), Dictionary of Medicines and Devices (UK), Drug Identification Number (Canada)).
b. The Diagnosis table, the Procedure table, and the Laboratory Result table have each been updated to allow for additional code types (e.g., ICD-10-CA (Canada), SNOMED CT (UK)).
c. In the Demographic table, Zip Code has been changed to Postal Code to enable various types of geographic identifiers.
d. For specific international Data Partners, the Enrollment table can now accept "Ambulatory Only" medical coverage.

5. What structures were introduced to increase run-time efficiencies in SCDM v8.0.0?
a. A required sort order was set for all tables, reducing processing need.
b. Identification variables (e.g., PatID) were set to numeric data type, reducing table size.
c. Variable length requirements for utilization codes were set to "minimal length necessary to contain code values," allowing for lengths specific to Data Partners and Datasets generated from Extract, Transform, Load procedures (often referred to as "ETLs"), and reduced table size.
d. Relocation of the Provider and Facility_Location variables from the SCDM Encounter table to newly created Provider and Facility tables reduced redundant information across multiple rows, and therefore reduced overall table size.

6. What were the results of these changes to run-time efficiencies?
Changes were assessed for disk space savings in the SCDM databases as well as execution run times in both the Quality Assurance and the Query Request Packages. Positive percents are improvements in efficiency (i.e., run-time decreases) and negative percents are reductions in efficiency (i.e., run-time increases). We found the following:
a. SCDM Database Disk Space: Overall, among the three Data Partners in which disk space utilization was observed (Humana, Marshfield, and internal MarketScan), savings in disk space ranged from 31% to 56%.
b. QA Package Run-times: Overall, among the four Data Partners assessed (CMS, Humana, Marshfield, and MarketScan), we observed a range of improvements in execution run times from -7.6% to 64.8%. Importantly, the CMS database, which contributes the largest volume of data in the Sentinel network, saw the greatest improvement in QA package run time, at 64.8%. The comparison is particularly meaningful because the CMS v7 and v8 databases had identical date ranges of source data, while the other Data Partners in the comparisons had different date ranges. Within the individual modules of the QA package (e.g., checking individual SCDM tables or different QA Level checks), the range was much greater, ranging from -87.5% to 92.4%.
c. QRP Package Run-times: We assessed run-time efficiencies for three Data Partners (Humana, Marshfield, and MarketScan). As the QRP Package itself was enhanced to improve run-time efficiencies across the time points when assessments were performed, the "Freeze Data" component of the QRP package enables a run-time comparison across the v7 versus v8 databases with the least amount of confounding. Here we observed improvements of 5.7% to 24%.

7. SCDM v8.0.0 was released December 10, 2020, so when are the data queryable?

The model release is one of several steps toward creating a queryable v8.0.0 Sentinel Distributed Database (SDD). These steps include:

  a. Creating v8.0.0-compliant ARIA and Data Characterization tools;
  b. Guiding Data Partners through transforming source data to v8.0.0-compliant data;
  c. Requiring new queries to be executed on v8.0.0-compliant versions of ARIA tools.

Access to the new Prescribing table and the new Provider Specialty variable will be via ad hoc queries and custom programming. Creating new ARIA tools and capabilities to <u>routinely utilize</u> the additional data will require new scopes of work.

8. What are the differences between Prescribing data and Dispensing data?

Prescribing data represent the prescription order written by a provider for a medication or medical product. At this time, in the U.S., only outpatient prescriptions that are filled and billed through insurance claims will appear in the Dispensing data table. Data Partners expected to populate the Prescribing table include Integrated Delivery System (IDS) partners as well as international partners.

9. Some individual providers have more than one medical specialty. Are all medical specialties for providers available in this new model?

When specialty information is available at a Data Partner, the model allows capture of only one medical specialty for each individual provider. Note also that:

a. Some Data Partners capture only one medical specialty per provider, while others sometimes capture multiple specialties in their source data.
b. Across the Sentinel network, there is limited ability on the part of Data Partners to determine, even when a provider has more than one specialty, under which specialty they were practicing during a patient's medical encounter.
c. Data Partners were instructed, when they have multiple specialties, to select the single specialty under which that provider practices most often.

10. Does this new model enable tracking patients and encounters from one ETL to another?

The model itself does not contain any information for this purpose. However, Data Partners have been instructed to establish systems to identify the same patients and encounters across multiple ETLs, whether that is within exclusively version 7 ETLs, version 8 ETLs, or between version 7 and version 8 ETLs.