



EVIDENCE NOW SOUTHWEST TECHNICAL SET-UP GUIDE

LICENSE

The following document is a derivative of the SAFTINet OMOP CDMV4 ETL Specification Document, found at <http://omop.org/CDM>

OMOP material is licensed under the Apache License Version 2.0, making it available to the public as Open Source with minimal restrictions. To learn more about what you need to do to use the CDM, please **review the License**.

Any redistributions of this work or any derivative work or modification based on this work should be accompanied by the following source attribution: "This work is based on work by the Observational Medical Outcomes Partnership (OMOP) and used under license from the RUF at <http://www.omop.org/publiclicense>

Any scientific publication that is based on this work should include a reference to <http://www.omop.org>.

This document was created specifically for Evidence Now Southwest.

Introduction

Thank you for choosing to participate in the Evidence Now Southwest (ENSW) project. The purpose of this guide is to assist you through the Technical Implementation process.

Your Data

During this process, patient-level data will need to be extracted from your EHR and provided to DARTNet. Your data file will be transferred to a secured portal where DARTNet will retrieve it. DARTNet will standardize and synthesize the data in order to update the Practice Performance Registry as specified by ENSW.

The initial data extraction will include up to 3 years of historical data elements. Data will be further de-identified and aggregated with other organizations and practices. On a quarterly basis, Performance Reports will be refreshed to include data submitted in the previous quarter.

Practice Performance Registry

Performance Reports will be published quarterly to a secure website managed by DARTNet. The reports compare data at an organization and Care Site (practice location) level. The names of organizations and practices who demonstrate the Top 5 Best Practices for the cohort are published for all users with access to the website to view. Your organization/practice name will only be visible to you and will remain anonymous to others unless your organization is identified as a Top 5 best practice. *Provider level reports under consideration.*

It is important to understand that if your Organization does not have data relevant to a given performance metric then your Organization/Care Site data may be excluded from that report. Users may still view the report but would not see your organization/practice listed. For example: a pediatric practice is submitting data. The cohort is for ages 18 and older who have diabetes and are seen only twice in the current year. The pediatric practice may have limited or no data, due to the age criteria, which would qualify them for the cohort. If they do see patients over the age of 18, they may have few who qualify based on the number of visits in the year. In this case, the pediatric practice data would be excluded from the cohort but saved for use in a more data relevant study. Users associated to the practice would only be able to view the report but would not see their Organization name displayed.

Data Transfer Set Up

DARTNet's Egnyte cloud storage server will be used to transfer data from the clinic to DARTNet. A link will be provided to each clinic that will lead to the specific directory for that clinic. Files are to be uploaded to this specific directory for processing.

Flat File Format

The basic file should be a .txt style text file with columns arranged in the order listed in this document. Individual column values should be separated by a pipe '|' character. A total of 9 files should be loaded in the initial round, one for each table in the table definition section.

Example: 1 row from a sample Organization file

This record (from Section 4.2):

Organization Table	
organization_source_value ¹	UC Internal Medicine
x_data_source_type	EHR
place_of_service_source_value	Academic Practice
organization_address_1	13199 E Montview Blvd
organization_address_2	Suite 300, Mail Stop F443
organization_city	Aurora
organization_state	CO
organization_zip	80045
organization_county	Arapahoe

Should be represented as follows in the file (the actual text should be all on one line):

UC Internal Medicine|EHR|Academic Practice|13199 E Montview Blvd|Suite 300, Mail Stop F443|Auora|CO|80045|Arapahoe

Users should apply the following rules when generating flat files

For further clarification on these rules and the format in general, see

http://en.wikipedia.org/wiki/Commaseparated_values#Basic_rules_and_examples

- Send a separate file for each data table
- When saving the files, use UTF-8 encoding. Files using Unicode or UTF-16 encoding are not compatible with the flat file processor.
- Files should be named using the following convention [table name].txt. Any spaces in the table name should be replaced by the underscore (_) character, so the file for the Drug Exposure table should be drug_exposure.txt
- Column values should be separated with the | character used as a delimiter

- Files should contain one record per row. No header row is needed, the first row should be actual data
- When double quotes (") occur within a string the entire string must be enclosed in double quotes, and the occurrence of the double quote itself must be duplicated. (Ex. John "No Name" Doe becomes "John ""No Name"" Doe"). This applies even when the double quotes surround the value (Ex. "John Doe" becomes ""John Doe"")
- Column values which contain the | delimiter should be enclosed in double quotes (Ex. This | That becomes "This | That")
- Datetime values should be in the following format 2012-01-09T24:00:00Z (example: 2012-01-09 16:15:00Z) and dates should be use the following format YYYY-MM-DD
- No column value should contain a 'newline' character, this will cause the parser to fail

Extraction Cohort Definition

For each data element, we ask for 3 years of historical data. The initial data extract should include all patients who had a visit in the given time period and were over 18 at the time of the visit.

The list of specific data elements to be included for this cohort of patients is specified below.

Patient-Level Data Elements Used to Create the Registry

Data Element	Direct Data Extraction to DARTNet	Destination Table
Patient Linking number	X	Found throughout all clinical data tables
Date of birth	X	x_demographic
Race	X	x_demographic
Ethnicity	X	x_demographic
Visits	X	visit_occurrence
CPT codes linked to each visit	X	procedure_occurrence
Diagnostic codes linked to each visit	X	condition_occurrence
Problem list	X	condition_occurrence
Medications	X	drug_exposure
Blood pressure - systolic	X	observation
Blood pressure- diastolic	X	observation
Weight	O	observation
Height	O	observation
Smoking status	X	observation
Smoking cessation referrals/treatment plan	X	Observation / procedure
LDL	X	observation
Total cholesterol	X	observation
HDL-C	X	observation
A1C	O	observation

How to use this section

In the tables below, The Source Field and Applied Rule fields are left blank for the practice to fill in. The source field should be filled in with the equivalent field in the source data. The Applied Rule field should contain any specialized rules (i.e. filtering, translation, combination of categories etc...) that are implemented when filling in the field.

Table Definitions

For complete table specifications, please see the accompanying Excel document.

Organization

The Organization table is the highest level of the care infrastructure hierarchy. Each organization may have multiple care sites. Providers can work at one or more care sites. Address information submitted with the organization will be used to create a new location record which will be linked to the organization record via the Location_ID field. During standardization, records that have the same value in all columns are considered duplicates and are combined into a single record.

Care_site

The Care Site table refers to the lower level of the provider care hierarchy. Individual provider care locations will be stored in this table. Records that have the same value in all columns are considered duplicates and are combined into a single record.

Provider

The Provider table contains information on local care providers including type and specialty. Providers are assigned to an individual care site. When passing through ROSITA to the grid, records with the same provider_source_value are assumed to be duplicates. If they differ in the other column values (i.e. last name), the first values to occur in the data file are used.

X_demographic

The X_Demographic table stores information about individual patients, the PHI elements of this record will be stripped out in the transformation to the analytic model. Address information will be limited and used to create a new location record. When passing through transformation to the analytic data set, records with the same person_source_value are assumed to be duplicates. If they differ in other column values (i.e. last name), the first values to occur in the data file are used. **While care_site_source_value and provider_source_value are not required, records in this table will be of much greater use if these field values can be provided.**

For performance registry purposes, direct patient identifiers are not needed. Those fields should be set to NULL in the x_demographic table as described in the table below.

Visit_occurrence

The Visit Occurrence table contains a record for each patient-provider encounter. The provider, patient and location are all stored as well as the type of visit. When passing through ROSITA to the grid, records that have the same value in all columns are combined into a single record. While

care_site_source_value is not required, records in this table will be of much greater use if the field value can be provided.

Drug_exposure

The Drug Occurrence table contains a record for each prescribed medication. The prescriber, patient, and prescription information are all stored as well as the associated visit and condition. When passing through ROSITA to the grid, records that have the same value in all columns are combined into a single record.

Condition_occurrence

The Condition Occurrence table contains a record for each patient condition. The codes associated with the conditions as well as the associated person, provider, and visits/encounters are also recorded. When passing through ROSITA to the grid, records that have the same value in all columns (with the exception of condition_occurrence_source_identifier) are combined into a single record. If two records have two different values for condition_occurrence_source_identifier but are the same in all other columns, they are assumed to refer to the same diagnosis and are combined into a single record.

Procedure_occurrence

The Procedure Occurrence table contains a record for each procedure. The type of procedure as well as the associated person and visit are recorded. When passing through ROSITA to the grid, records that have the same value in all columns (with the exception of procedure_occurrence_source_identifier) are combined into a single record. If two records have two different values for procedure_occurrence_source_identifier but are the same in all other columns, they are assumed to refer to the same procedure and are combined into a single record.

Observation

The Observation table contains records for labs, measurements such as height and weight, etc... It is also where information from Past Medical History, Past Surgical History, Allergy, and Social/Personal History are stored. During processing, records that have the same value in all columns (with the exception of observation_source_identifier) are combined into a single record. If two records have two different values for observation_source_identifier but are the same in all other columns, they are assumed to refer to the same observation and are combined into a single record.