ESP: MDPHnet Data Model

High Level Specification

Version 1.1

July 2013

Contents

Revision History

Version Number	Modification Date	By	Description of Changes
0.3	13 March 2012	R Schaaf	Initial version
1.0	09 July 2013	B Zambarano	Production release
1.1	16 July 2013	J Brown, M Davies, A Leibowitz	Text edits for release to Data Partners. Posted to query tool

1 Introduction

The EHR Support for Public Health (ESP) is a disease surveillance software application that can extract and analyze data from electronic health record systems for events of public health importance. ESP organizes raw data extracted from EHR systems, maps them to disease categories, analyzes these data for conditions of public health interest, and transmits either case-level or aggregate data to public health agencies

This design note describes the data model used by MDPHNet to enable distributed querying. This data model - referred to as ESP: MDPHNet - is a simplified version of the full ESP data model and is generated via SQL scripts that execute against the full ESP data model. The ESP: MDPHNet data model is motivated by a desire to:

- Provide a mapping layer between the full set of ESP tables and the MDPHNet application to allow the organization of ESP data to evolve over time without adversely impacting MDPHNet.

- Adapt the ESP data model to provide a set of database tables that are better suited to the needs of the MDPHNet query application.

- Improve database query performance relative to what would have been achievable running queries against the full set of ESP tables.

2 Data Model Components

The simplified data model includes three sets of tables:

- Patient-level and encounter-level tables. Currently, these include demographic, encounter, diagnosis and detected disease data tables. Additional tables can be added in later project phases.
- Unique value tables (UVTs). These are used to efficiently gather lists of categorical (i.e., ”pick list”) information.
- Summary tables. These provide counts of visits and patients stratified by center, age group, time period, encounter type and diagnosis. The summary tables can be used to facilitate querying and query response by avoiding access to patient or encounter level data.

Tables may include B-tree indexes which organize records to improve search efficiency.

3 Patient and Encounter Level Tables

The ESP: MDPHNet data model includes the following tables.

Whenever possible, the organization and format of information in these tables is patterned after the Mini-Sentinel Common Data Model (MSCDM), v2.1. Use of the MSCDM will facilitate cross-project sharing of querying capabilities.

- **ESP_DEMOGRAPHIC**: Contains one record per patient with the most recent information on birth date, sex, ethnicity, and race.

- **ESP_ENCOUNTER**: Contains one record per patient visit. This table includes provider, visit date, visit location, encounter type and age group information.

- **ESP_DIAGNOSIS**: Contains one record per combination of patient, visit and diagnosis.

- **ESP_DISEASE**: Contains one record per combination of patient and detected disease\case based on the ESP case detection algorithms. The table includes detection criteria, review workflow status, reviewer notes and derived variables for 5 year, 10 year and Mini-Sentinel age groups.

3.1 ESP_DEMOGRAPHIC

This table contains the most recent patient demographic information on birth date, sex and race.

Column	Data Type	Description
centerid	character varying(128)	Center identifier.
patid	character varying(128)	Arbitrary person-level identifier. Used to link across tables.
birth_date	integer	Date of birth, formatted as a SAS Date (number of days since Jan 1, 1960).
sex	character(1)	Values returned: M=Male F=Female U=Unknown.
hispanic	character(1)	A person of Hispanic race. Values returned: Y=Yes, N=No, U=Unknown.
race	integer	Values returned: 0=Unknown 1=American Indian or Alaska Native 2=Asian 3=black or African American 4=Native Hawaiian or Other Pacific Islander 5=White
zip5	character varying(5)	5 character zip code for patient's primary address.

PATID is used as a primary key constraint

There are seven B-tree indexes on the table, one for each field.

CENTERID, RACE and SEX are constrained by foreign key constraints against their respective UVT tables described below.

3.2 ESP_ENCOUNTER

This table contains patient visit information along with derived variables for 5 year, 10 year and Mini-Sentinel age groups.

Column	Data Type	Description
centerid	character varying(128)	Center identifier
patid	character varying(128)	Arbitrary person-level identifier. Used to link across tables.
encountered	character varying(128)	Arbitrary encounter-level identifier. Used to link across the Encounter and Diagnosis tables.
a_date	integer	Encounter date, formatted as s SAS Date (number of days since Jan 1, 1960).
d_date	integer	Date the encounter was closed, formatted as a SAS Date (number of days since Jan 1, 1960).
provider	character varying(128)	Provider code for the provider who is most responsible for this encounter.
facility_location	character varying(100)	Facility name.
enc_type	character varying(10)	Encounter type.
facility_code	character varying(30)	Local facility code that identifies hospital or clinic.
enc_year	integer	Calendar year for this encounter date.
age_at_enc_year	integer	Patient age as of Jan 1 of the year of the encounter.
age_group_5yr	character varying(5)	Patient age group (specified in 5 year intervals) as of Jan 1 of the year of the encounter.
age_group_10yr	character varying(5)	Patient age group (specified in 10 year intervals) as of Jan 1 of the year of the encounter.
age_group_ms	character varying(5)	Patient age group (using Mini-Sentinel age groups) as of Jan 1 of the year of the encounter.

ENCOUNTERID is used as a primary key constraint

There are fourteen B-tree indexes on the table, one for each field.

AGE_GROUP_5YR, AGE_GROUP_10YR, AGE_GROUP_MS, ENC_TYPE, ENC_YEAR, FACILITY_CODE, and PROVIDER are constrained by foreign key constraints against their respective UVT tables described below.

PATID is constrained by foreign key constraint against ESP_DEMOGRAPHIC.PATID

3.3 ESP_DIAGNOSIS

This table contains patient diagnosis information for each encounter. For ease of querying, this table also incorporates information from the visit associated with each diagnosis and also 3-digit and 4-digit substrings of the recorded diagnosis code (variable: dx).

Column	Data Type	Description
centerid	character varying(128)	Center identifier.
patid	character varying(128)	Arbitrary person-level identifier. Used to link across tables.
encountered	character varying(128)	Arbitrary encounter-level identifier. Used to link across the Encounter and Diagnosis tables.
a_date	integer	Encounter date, formatted as s SAS Date (number of days since Jan 1, 1960).
provider	character varying(128)	Provider code for the provider who is most responsible for this encounter.
enc_type	character varying(10)	Encounter type.
dx	character varying(10)	ICD-9 diagnosis code.
dx_code3dig	character varying(4)	3-digit ICD-9 code.
dx_code4dig	character varying(5)	4-digit ICD-9 code, with decimal point removed.
dx_code4dig_with_dec	character varying(6)	4-digit ICD-9 code, with decimal point.
dx_code5dig	character varying(6)	5-digit ICD-9 code, with decimal point removed.
dx_code5dig_with_dec	character varying(7)	5-digit ICD-9 code, with decimal point.
facility_location	character varying(100)	Facility name.
facility_code	character varying(30)	Local facility code that identifies hospital or clinic.
enc_year	integer	Calendar year for this encounter date.
age_at_enc_year	integer	Patient age as of Jan 1 of the year of the encounter.
age_group_5yr	character varying(5)	Patient age group (specified in 5 year intervals) as of Jan 1 of the year of the encounter.
age_group_10yr	character varying(5)	Patient age group (specified in 10 year intervals) as of Jan 1 of the year of the encounter.
age_group_ms	character varying(5)	Patient age group (using Mini-Sentinel age groups) as of Jan 1 of the year of the encounter.

N.B.: DX, DX_CODE_4DIG, DX_CODE_4DIG_WITH_DEC, DX_CODE_5DIG, and DX_CODE_5DIG_WITH_DEC may include codes with less than 4 or 5 digits, if that reflects the corresponding code in the medical record.

PATID, ENCOUNTERID, DX comprise the primary key constraint There are nineteen B-tree indexes on the table, one for each field.

There are two B-tree varchar_pattern_ops indexes on the table one for DX and another for DX_CODE_5DIG.

DX, DX_CODE_3DIG, DX_CODE_4DIG_WITH_DEC, and DX_CODE_5DIG_WITH_DEC are constrained by foreign key constraints against their respective UVT tables described below.

PATID is constrained by foreign key constraint against ESP_DEMOGRAPHIC.PATID.

ENCOUNTERID is constrained by foreign key constraint against ESP_ENCOUNTER.ENCOUNTERID.

3.4 ESP_DISEASE

This table contains one record per combination of patient and detected disease along with the detection criteria, review workflow status, reviewer notes and derived variables for 5 year, 10 year and Mini-Sentinel age groups. This table is generated via execution of the various ESP disease detection algorithms.

Column	Data Type	Description
centerid	character varying(128)	Center identifier.
patid	character varying(128)	Arbitrary person-level identifier. Used to link across tables.
condition	character varying(100)	Condition that triggered the disease detection.
date	integer	Date on which the disease was detectable, formatted as a SAS Date (number of days since Jan 1, 1960).
age_at_detected_year	integer	Patient age as of January 1 of the year of the disease was detected.
age_group_5yr	character varying(5)	Patient age group (specified in 5 year intervals) as of Jan 1 of the year of the encounter.
age_group_10yr	character varying(5)	Patient age group (specified in 10 year intervals) as of Jan 1 of the year of the encounter.
age_group_ms	character varying(5)	Patient age group (using Mini-Sentinel age groups) as of Jan 1 of the year of the encounter.
criteria	character varying(255)	Detection criteria.
status	character varying(32)	Review workflow status. One of: AR – Awaiting review UR – Under review RM – Reviewed by MD FP – False positive Q – Confirmed case S – Transmitted to health department
notes	text	Review notes.

PATID, CONDITION, DATE comprise the primary key constraint

There are ten B-tree indexes on the table, one for each field except NOTES.

CONDITION, CRITERIA and STATUS are constrained by foreign key constraints against their respective UVT tables described below.

PATID is constrained by foreign key constraint against ESP_DEMOGRAPHIC.PATID.

4 Unique Value Tables

The data model includes a set of unique value tables (UVTs) as a convenience for the application to use in populating "pick lists" of categorical information. The format of the UVT tables is generally:

Column	Data Type	Description
item_code	depends on the data item	Data value.
item_text	character varying	Display text corresponding to item_code.

In addition, the two UVTs for DX_4DIG and DX_5DIG have an additional field called "item_code_with_dec", which includes the ICD9 code with the decimal point.

UVT tables are provided for the follow items.

UTV Table	Associated Data Elements
UTV_SEX	esp_demographic.sex
UTV_RACE	esp_demographic.race
UTV_PROVIDER	esp_encounter.provider
UTV_CENTER	esp_demographic.centerid
UTV_SITE	UTV **item_code** is associated with: esp_encounter.facility_cide UTV **item_text** is associated with: esp_encounter.facility_location
UTV_PERIOD	esp_encounter.enc_year
UTV_ENCOUNTER	esp_encounter.enc_type
UTV_AGEGROUP_5YR	esp_encounter.age_group_5yr
UTV_AGEGROUP_10YR	esp_encounter.age_group_10yr
UTV_AGEGROUP_MS	esp_encounter.age_group_ms
UTV_DX	esp_diagnosis.dx
UTV_DX_3DIG	esp_diagnosis.dx_code_3dig
UTV_DX_4DIG	esp_diagnosis.dx_code_4dig, dx_code_4dig_with_dec
UTV_DX_5DIG	esp_diagnosis.dx_code_5dig, dx_code_5dig_with_dec
UTV_DETECTED_CONDITION	esp_disease.condition
UTV_DETECTED_CRITERIA	esp_disease.criteria
UTV_DETECTEDSTATUS	esp_disease.status
UTV_ZIP5	es_demographic.zip5

5 Summary Tables

The data model includes summary tables that provide counts of visits and patients stratified by center, age group, period, encounter type and diagnosis.

In addition, there is a table summarizing Influenza-like Illness (ILI) cases that conforms to CDC reporting standards. ILI case detection is based on the ESP ILI case algorithm.

5.1 ESP_DIAGNOSIS_ICD9_3DIG

The ESP_DIAGNOSIS_ICD9_3DIG table is analogous to the "3-Digit ICD-9 Diagnosis Summary Table Structure" defined as part of the Mini-Sentinel Common Data Model (v2.1).

This table provides a count of unique individuals stratified by age group, sex, year, and care setting who had at least one encounter recorded with the diagnosis of interest based on the first 3 digits of the diagnosis code. Three different sets of age groupings are represented in the summary table. The different age groupings are distinguished by "age_group_type".

Column	Data Type	Description
centerid	character varying(128)	Center identifier.
age_group_type	character varying(15)	Distinguishes between the following types of age groups: 'Age Group 5yr', 'Age Group 10yr', Age_Group_MS'.
ageo_group	character varying(5)	The following age groups are used: Age Group 5yr: 0-4, 5-9, 10-14, 15-19, 20-24, …, 100+ Age Group 10yr: 0-9, 10-19, 20-29, 30-39, …, 100+ Age Group MS: 0-1, 2-4, 5-9, 10-14, 15-18, 19-21, 22-44, 45-64, 65-74, 75+
sex	character varying(1)	Sex: 'M', 'F', 'U'
period	integer	Time period represented as a 4 digit calendar year.
code_	character varying(4)	3 digit ICD9 diagnosis code.
setting	character varying(10)	Encounter type.
members	bigint	Count of members who had one or more encounters during the period with the diagnosis of interest recorded.
events	bigint	Count of encounters (events) observed with the diagnosis of interest recorded.
dx_name	character varying(150)	Diagnosis name.

There are seven B-tree indexes on the table, one for each field except MEMBERS, EVENTS and DX_NAME.

5.2 ESP_DIAGNOSIS_ICD9-4DIG

The ESP_DIAGNOSIS_ICD9_4DIG table is analogous to the "4-Digit ICD-9 Diagnosis Summary Table Structure" defined as part of the Mini-Sentinel Common Data Model (v2.1).

This table provides a count of unique individuals stratified by age group, sex, year, and care setting who had at least one encounter recorded with the diagnosis of interest based on the first 4 digits of the diagnosis code. Three different sets of age groupings are represented in the summary table. The different age groupings are distinguished by "age_group_type".

Note: The summary table may also include 3 digit codes for those diagnoses that were entered using a 3-digit code.

Column	Data Type	Description
centerid	character varying(128)	Center identifier.
age_group_type	character varying(15)	Distinguishes between the following types of age groups: 'Age Group 5yr', 'Age Group 10yr', Age_Group_MS'
ageo_group	character varying(5)	The following age groups are used: Age Group 5yr: 0-4, 5-9, 10-14, 15-19, 20-24, …, 100+ Age Group 10yr: 0-9, 10-19, 20-29, 30-39, …, 100+ Age Group MS: 0-1, 2-4, 5-9, 10-14, 15-18, 19-21, 22-44, 45-64, 65-74, 75+
sex	character varying(1)	Sex: 'M', 'F', 'U'
period	integer	Time period represented as a 4 digit calendar year.
code_	character varying(5)	4 digit ICD9 diagnosis code. *Note: This field may contain 3 digit code if that was entered as the diagnosis.*
setting	character varying(10)	Encounter type.
members	bigint	Count of members who had one or more encounters during the period with the diagnosis of interest recorded.
events	bigint	Count of encounters (events) observed with the diagnosis of interest recorded.
dx_name	character varying(150)	Diagnosis name.

There are seven B-tree indexes on the table, one for each field except MEMBERS, EVENTS and DX_NAME.

5.3 ESP_DIAGNOSIS_ICD9-5DIG

The ESP_DIAGNOSIS_ICD9_5DIG table is analogous to the "5-Digit ICD-9 Diagnosis Summary Table Structure" defined as part of the Mini-Sentinel Common Data Model (v2.1).

This table provides a count of unique individuals stratified by age group, sex, year, and care setting who had at least one encounter recorded with the diagnosis of interest based on the first 5 digits of the diagnosis code. Three different sets of age groupings are represented in the summary table. The different age groupings are distinguished by "age_group_type".

Note: The summary table may also include 3 or 4 digit codes for those diagnoses that were entered using a 3 or 4-digit code.

Column	Data Type	Description
centerid	character varying(128)	Center identifier.
age_group_type	character varying(15)	Distinguishes between the following types of age groups: 'Age Group 5yr', 'Age Group 10yr', Age_Group_MS'
ageo_group	character varying(5)	The following age groups are used: Age Group 5yr: 0-4, 5-9, 10-14, 15-19, 20-24, …, 100+ Age Group 10yr: 0-9, 10-19, 20-29, 30-39, …, 100+ Age Group MS: 0-1, 2-4, 5-9, 10-14, 15-18, 19-21, 22-44, 45-64, 65-74, 75+
sex	character varying(1)	Sex: 'M', 'F', 'U'
period	integer	Time period represented as a 4 digit calendar year.
code_	character varying(6)	4 digit ICD9 diagnosis code. *Note: This field may contain 3 or 4 digit code if that was entered as the diagnosis.*
setting	character varying(10)	Encounter type.
members	bigint	Count of members who had one or more encounters during the period with the diagnosis of interest recorded.
events	bigint	Count of encounters (events) observed with the diagnosis of interest recorded.
dx_name	character varying(150)	Diagnosis name.

There are seven B-tree indexes on the table, one for each field except MEMBERS, EVENTS and DX_NAME.

5.4 ILI_SUMMARY

The ILI_SUMMARY table is used to generate a set of standard reports for ILI syndrome surveillance. A row represents a weekly summary for a given age group at a given healthcare site.

Column	Data Type	Description
age_group	text	Age groups in years 0-4, 5-24, 25-49, 50-64, 65+
period_end	date	For each week, date as of Saturday.
week	character varying(5)	MMWR Week value (see http://wwwn.cdc.gov/nndss/document/MMWR_Week_overview.pdf).
zip5	character varying(5)	Five character zip code for healthcare site where diagnosis was made.
center	character varying(100)	Name of healthcare site where diagnosis was made.
cdc_site_id	character varying(50)	Site ID as provided by CDC for reporting purposes.
ili_counts	bigint	Count of ILI cases detected.
tot_counts	integer	Count of total patient encounters.