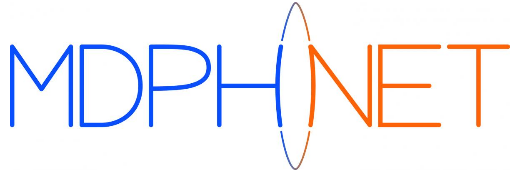
****

**ESP: MDPHnet Data Model**

**High Level Specification**

Version 1.1

July 2013

Contents

[1 Introduction 1](#_Toc361753394)

[2 Data Model Components 2](#_Toc361753395)

[3 Patient and Encounter Level Tables 3](#_Toc361753396)

[3.1 ESP\_DEMOGRAPHIC 4](#_Toc361753397)

[3.2 ESP\_ENCOUNTER 5](#_Toc361753398)

[3.3 ESP\_DIAGNOSIS 6](#_Toc361753399)

[3.4 ESP\_DISEASE 8](#_Toc361753400)

[4 Unique Value Tables 9](#_Toc361753401)

[5 Summary Tables 10](#_Toc361753402)

[5.1 ESP\_DIAGNOSIS\_ICD9\_3DIG 11](#_Toc361753403)

[5.2 ESP\_DIAGNOSIS\_ICD9-4DIG 12](#_Toc361753404)

[5.3 ESP\_DIAGNOSIS\_ICD9-5DIG 13](#_Toc361753405)

[5.4 ILI\_SUMMARY 14](#_Toc361753406)

**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. |