Report on Optimal Installation of ESP-VAERS

By Bob Zambarano and Chris Wright, 31 May 2019.

Contents

[Overview 2](#_Toc10191380)

[Section 1: Building the Extraction-Transformation-Load (ETL) system. 2](#_Toc10191381)

[Understanding the structure and content requirements of the ESP data load files. 3](#_Toc10191382)

[Determining the source data elements from the clinical practice EHR that correspond to the target data elements in the ESP data load files. 3](#_Toc10191383)

[Writing the code to extract the data from the clinical practice EHR and transform the data into the ESP data load file structure. 4](#_Toc10191384)

[Determining how the nightly ETL process will be run, and how the ESP data load files will be made available to the ESP load system. 4](#_Toc10191385)

[Section 2: Installation and Configuration of ESP VAERS 5](#_Toc10191386)

[Determining the storage and processing requirements of the ESP server 5](#_Toc10191387)

[Installing the server into the network enclave or data center 6](#_Toc10191388)

[Install ESP 6](#_Toc10191389)

[Load data from EHR system 7](#_Toc10191390)

[Configure ESP data mappings for Labs and Immunizations 7](#_Toc10191391)

[Section 3: Running VAE detection and VAE\_Listing for review of configuration 10](#_Toc10191392)

[Running VAE detection 10](#_Toc10191393)

[Running the VAE listing 10](#_Toc10191394)

[Section 4: MDM-T02 message interface 11](#_Toc10191395)

[Section 5: Setting up PHINMS and sending VAERS messages. 11](#_Toc10191396)

## Overview

ESP-VAERS is a software system for detecting vaccine adverse events (VAEs) and reporting those events to public health authorities using the VAERS reporting structure.

The system is installed as follows:

1. A data extraction and transfer process must be developed to move data from the clinical EHR system to ESP. (See Section 1 below: Building the ETL system).
2. An ESP server must be set up with ESP installed, the data loaded and configured (See Section 2 below: Installation and Configuration of ESP VAERS).
3. An initial VAE detection run must be completed, and the VAE\_Listing generated for review and verification that VAEs are being correctly detected. Updates to the system configuration may be indicated (See Section 3 below: Running VAE detection and VAE\_Listing for review of configuration).
4. The potential VAE cases will be used to create messages for clinical review. The messages are generated in HL7 2.3.1 MDM-T02 format. The EHR system must have an interface to accept these messages, which become part of the patient medical record and which provide a link back to an ESP web page for confirmation or rejection. There are two types of EHR interface messages: messages indicating a potential VAE, and messages indicated a confirmed VAE and VAERS report sent (See Section 4 below: MDM-T02 message interface).
5. The confirmed VAE cases are used to create VAERS messages for transfer to the PHIN-MS secure message transfer service to the CDC’s electronic VAERS message receiving address as managed by GDIT (see Section 5 below: Setting up PHINMS and sending VAERS messages).

## Section 1: Building the Extraction-Transformation-Load (ETL) system.

ETL is the process for moving data from one data processing system to another. For ESP, this is the process of copying data from the clinical practice EHR system to the ESP data warehouse. For any site installing ESP-VAERS, the first step of the process is understanding the data requirements for the ESP system. ESP is an enterprise-level data warehouse-based application, which is provisioned nightly with all new and updated patient clinical data available in the clinical practice EHR system. There is a common misconception that ESP searches the clinical practice EHR system for VAEs, then extracts the necessary data for reporting, but this is not the case. All relevant patient data is copied to ESP, where is can be searched, VAE cases identified, and VAERS reports compiled. The VAERS reporting form requires PHI data, so ESP data provisioning requirements include all new and updated:

* Patient demographic information such as name, address, phone number, race, date of birth.
* Health care provider information such as name, address, phone number, care facility.
* Vaccination information such as date and time of vaccination, vaccine manufacturer, lot number, vaccine body site.
* Prescription information such as prescription name, order date, start date.
* Lab results data such as test name, order date, result date, results.
* Clinical Diagnosis data such as diagnosis code (ICD10), date of diagnosis.
* Clinical problems such as diagnosis codes and date of problem recording.

The ESP ETL system is built and installed according to the following steps:

1. Understanding the structure and content requirements of the ESP data load files.
2. Determining the source data elements from the clinical practice EHR that correspond to the target data elements in the ESP data load files.
3. Writing the code to extract the data from the clinical practice EHR and transform the data into the ESP data load file structure.
4. Determining how the nightly ETL process will be run, and how the ESP data load files will be made available to the ESP load system.

We will review these steps in sequence. The first two steps can be completed prior to the ESP installation be done. The third step cannot be completed until an ESP instance is installed and ready for testing the data. The fourth step requires an ESP installation. ESP installation is described in the second section below.

### Understanding the structure and content requirements of the ESP data load files.

Currently, the ESP data warehouse is provisioned by generation of a set of delimited text files created from the clinical practice EHR[[1]](#footnote-1). These files are described in the accompanying document: “ESP\_Filespec\_v1.5\_VAERS”. This document is an Excel workbook. The workbook tabs are arranged in the optimal order of review to understand the ESP ETL file structure requirements. The first tab provides an overview of the Workbook contents and provides descriptions of columns of interest. The second tab describes the files to be produced, including the file naming requirements and the file column layouts. The subsequent tabs describe the data contents of each of the ETL delimited files.

The ESP system is used to support a number of disease reporting and surveillance activities, and the data requirements for those activities are more extensive than required for VAERS reporting. The data requirements for the ESP system must be based on the overall reporting and surveillance purposes of the ESP system, but if the system will only be used for VAERS reporting, the workbook tab for each file specification includes a column “VAERS Required”. These fields must be populated from the EHR to support VAERS reporting. Additional fields are optional; however, the developer must keep in mind that the order of delimited fields is critical, and if a field is omitted, delimiters must still be used to maintain the correct column count.

Delimited text files can be created by many data reporting systems. Any data reporting system that can produce delimited text files can be used to develop the ESP data provisioning interface. Most clinical practice EHR system support SQL-based reporting systems, and the ESP code repository includes a number of SQL data extract script samples that can be adapted to generate the daily ETL data. These samples include SQL for Epic Clarity, GE Centricity, and Cerner EHR systems. These can be modified to conform to any EHR sites needs and set up fairly easily. These are available for download from <https://gitlab.com/ESP-Project/esp_tools/tree/master/sample_etl>. Every clinical practice EHR is unique, and these sample scripts must be modified and tested for use at any site.

### Determining the source data elements from the clinical practice EHR that correspond to the target data elements in the ESP data load files.

For clinical practice sites using EHR systems for which sample extract scripts are not available, an extraction process must be developed for the ESP installation. This involves identifying how the corresponding data is stored in the clinical practice EHR, and determining how this data can be extracted into the data structure required for the delimited file. The ESP\_filespec document is a useful tool for designing and documenting this process. For each file to be developed, the workbook tab for that file should be used to determine the required field (VAERS Required), and additional columns can be added to the spreadsheet to specify the data source(s) in the clinical practice EHR. This document then becomes the ETL design specification for developing the ETL code.

### Writing the code to extract the data from the clinical practice EHR and transform the data into the ESP data load file structure.

The ETL process must meet the following basic requirements:

1. It must be able to extract all relevant patient clinical data as specified in the ESP\_filespec document.
2. It must be able to create file structures matching the forms specified in the ESP\_filespec document.
3. It must be able to generate these files for healthcare services provided on a specified date.
4. It must be able to run as a scheduled unattended process on a daily (or otherwise regular) basis.

The ETL code, once developed or modified from the available samples, must be tested. Depending on the site, a formal or informal testing plan should be developed. Tests should be performed to assess the following:

1. Does the ETL file format match the specification?
2. Can ESP load the data without errors?
3. Does the data provided correctly correspond to data in the clinical practice EHR for the specified extraction date?

### Determining how the nightly ETL process will be run, and how the ESP data load files will be made available to the ESP load system.

The ESP system must be installed in the same network enclave or data center as the source EHR. The extraction process generates the data files from the EHR and places them in the ESP incoming data folder. ESP loads all available data files once a day, or as often as data updates are provided.

Once setup, the extraction process runs nightly and provides the data interface between the EHR system and ESP VAERS. Each site will design and develop its own nightly scheduled ETL process, in order to meet site-specific policy and procedure requirements, and in order to work most efficiently with the clinical practice EHR. Here are actual use-case examples of how this can be done:

1. Direct pull from ESP server: Python scripts, with embedded SQL code are run as a scheduled job at a specific time each night on the ESP server. These scripts query the EHR for the required data, and generate the required files. The ETL files are written to a file folder on the ESP server. When the ETL scripts are complete, the scheduled job moves on to the next step of loading the data to ESP, then running the VAE detection algorithms, and so on.
2. Indirect pull from ESP server: MUMPS scripts are run against a Cache database from a scheduled job at a specific time each night from the EHR data server. The ETL files are created in a folder accessible via sFTP from the ESP server. A separate job on the ESP server checks this folder on a regular basis for new files, and moves them over to the ESP server when they are available. Once moved, the rest of the ESP load and detect processes are run.
3. Push to ESP: C# programs are created with embedded t-SQL to run data queries on a MS SQL-Server database containing clinical practice EHR data. These files are streamed to a shared filesystem that is available to both the ETL process running against the MS SQL-Server database, and ESP’s data load process. ESP regularly checks this filesystem folder for new files, and loads them when they are available.

The nightly data load and detection run on the ESP server is controlled by a shell script that is installed into the Linux crontab utility. An sample file is available at <https://gitlab.com/ESP-Project/esp_tools/blob/master/sample_etl/daily_batch.sh>. This file is a simple example of how to set up a scheduled load and detection process on the ESP side.

## Section 2: Installation and Configuration of ESP VAERS

The ESP system installation and configuration will follow these steps:

1. Determining the storage and processing requirements of the ESP server.
2. Obtain the server (physical or virtual) and install into the network enclave or data center.
3. Install ESP
4. Load data from EHR system
5. Configure ESP data mappings for Labs and Immunizations.

### Determining the storage, memory and processing requirements of the ESP server

Determination of the storage, memory and processing requirements for the ESP server must be done before a server is acquired and installed. There are two factors to take into account when determining these server size characteristics:

1. How much data will the EHR be providing on a nightly basis?
2. How long will EHR data be maintained in the ESP system (how much historic patient data will be maintained)?

To determine the first factor, it is necessary to query the clinical practice EHR to determine the basic counts of the following with respect to the ETL files created on a nightly basis for two sources:

* How many new or updated lab result file rows?
* How many new or updated encounter file rows?

These are typically among the largest files generated nightly, and the encounter and lab data is subjected to the most processing and review, and so these are the best indication of nightly data processing load requirements. With the information about data flow rates for these two files, server size can be calculated. A basic ESP system supporting a Postgres database and the Python processes that run the VAE detection process must have two CPUs (cores). From there:

* Add an additional CPU if you will be processing more than 40,000 encounter rows per day, and an additional CPU for each multiple of 40,000.
* Add an additional CPU if you will be processing more than 40,000 lab records per day, and an additional CPU for each multiple of 40,000.

For example, a system that processes an average of 20,000 encounters and 38,000 lab result rows per day will find that 2 CPUs (cores) are sufficient. A system with 75,000 encounters and 125,000 lab results per day would need at least 6 CPUs.

Storage requirements are a function of the number of patients under active care and the amount of historic patient data that the system will maintain. For ESP-VAERs reporting, the system must maintain at least a year of patient prior histories. Many of the VAE detection algorithms look back a year into the patients’ medical records for exclusionary events. In addition, for validation purposes, it is useful to maintain an additional year of patients’ medical histories so that test VAE cases can be created and reviewed when the system is installed. If the ESP installation is being used for other disease case reporting and surveillance purposes, additional years of patient histories will be required.

A simple heuristic for determining storage requirements for ESP-VAERS is based on the number of active patients and the number of years of patient medical histories the system will maintain. Active patients are patients with at least one encounter with a care provider in a year.

Count of active patients \* number of years \* 0.00025 = Gigabytes of storage required

For example, a system with 300,000 patients and 3 years of data would require:

 300,000\*3\*0.00025=225Gb storage

This estimate includes storage for the ESP relational database tables, compressed delimited text files, compressed database backup dumps, and 15% free space overhead. This assumes that only the required VAERS data fields are populated.

ESP stored data in a relational database management system (RDBMS), typically PostgreSQL. RDBMS performance is highly dependent on the amount of system memory available. The larger your RDBMS, the more memory your system can utilize to improve performance.

A rule-of-thumb we use is for each 50GB of storage used by the RDBMS, you should have 2GB of memory. Keep in mind that the storage heuristic above is for ALL storage plus overhead, not just the database. The RDBMS will take up about 50% of the used storage. Also keep in mind that a new ESP system will not have as much data stored as a machine that has been collecting and maintaining patient histories for several years. It may be appropriate to start out with a smaller amount of system memory, and expand this along with the size of your database.

### Installing the server into the network enclave or data center

Once the server size requirements have been obtained, an ESP server must be set up in the network enclave or data center. This may be a physical or virtual server. Basic server requirements are:

* An actively supported Linux distribution release.
* OpenSSH installed
* A static IP number assigned

### Install ESP

To complete an ESP installation, one follows the step-by-step instructions on the accompanying document “HowTo - Install and Configure ESP on Ubuntu 18.04”. These instructions are specific to Ubuntu 18.04, but they contain high-level instructions for installation on other Linux distributions as well. The installation follows these steps:

1. Create the ESP user and install the required software infrastructure.
2. Download the ESP software and run the installation
3. Create the ESP database and ESP database user
4. Create the filesystem locations required by ESP input and output processes.
5. Configure the Apache web server for the ESP administrative interface
6. Configure iptables for controlled access to the server
7. Configure the basic ESP system and create the admin user

### Load data from EHR system

The ETL system developer under Section1 above is now implemented for use with ESP. This includes extracting and loading patient histories (at least one year, optimally two years, optionally more), as well as setting up and running the automated nightly ETL process.

### Configure ESP data mappings for Labs and Immunizations

Once data is loaded to ESP, there are two final configuration steps which must be completed prior to running.

1. Mapping local lab test codes to ESP lab concepts
2. Mapping local vaccine codes to CDC standard vaccine codes

These mapping tasks must be completed after patients’ medical history data are loaded, and as part of system maintenance on a regular basis going forward. The mapping must be a standard maintenance task in order to detect and map any new labs and vaccines that appear in the medical records at a clinical practice.

A number of ESP VAEs are determined based on lab test results. Most health care organizations do not maintain lab test records uniformly coded to national standards. In order for the ESP to correctly assess a lab test results as pertaining to a VAE, the local lab test code and lab test name must be mapped to an ESP lab concept. ESP VAE detection does not require mapping of lab test codes and names to any standard code system.

Before lab test mapping can be performed, an ESP command must be run that populates the data table used to determine the set of available local labs. From the Linux command line, run the command:

[$ESPHOME]/bin/esp concordance

The concordance command must be run prior to using the lap mapping interface with the ESP system in order to detect any new lab tests in the medical record.

The lab test mapping process is performed via the ESP administrative interface, which provides an interactive “Unmapped Lab Tests Report”.



The report provides a simple dialog with a drop down pick list.



The user should pick one of the test name abbreviations on the list then select the button labelled “Search lab native names by condition”. If there a large number of unmapped labs, and the user picks “All conditions”, the interface will become very slow.

Below we have searched for “white”. The target was tests for White Blood Cell counts, but that is not what was turned up:



These codes may be permanently ignored so they don’t turn up in future mapping searches. Click on the box on the left side of each row and then click on the “Ignore Selected Codes” button.

Alternatively, if a test had been available for mapping, the link at the right side of the row is available for mapping the test to a lab concept. The lab mapping interface looks like so:



The only field that MUST be set is the value for “Test name:” on the lower left side in this view. The remain values have correct defaults.

Mapping vaccines uses a simpler interface.



Unmapped values in the local vaccine list are shown on the left, the list of mapped values is shown on the right. If a local vaccine value is not considered a vaccine, it should be left unmapped. Click on the “Edit” link to map an unmapped entry.



Use the drop-down list to pick a CDC standard vaccine name to map the local value to, then click save.

## Section 3: Running VAE detection and VAE\_Listing for review of configuration

Once the ETL has been developed and tested and the ESP system has been installed and configured and the ETL process has been used to load patient data, the next steps are to run VAE detection against the loaded data, and generate a VAE listing of all detected cases for review and validation.

### Running VAE detection

Running VAE detection requires submission of an ESP command from the Linux command line. The command does accept a number of arguments and requires at least one. The command syntax is:

[$ESPHOME]/ bin/esp vaers [options]

The options explained:

-b BEGIN\_DATE The start of VAE detection period

-e END\_DATE The end of the VAE detection period

-l Run Lab Results Heuristics

-d Run Diagnostics Heuristics

-p Run Prescription Heuristics

-r Run pRoblem Heuristics

-a Run All Heuristics

You must include one of l, d, p, r, or a (lab, diagnostic, prescription, problem or all heuristics).

For example:

[$ESPHOME]/bin/esp vaers -a -b 20180101

The above command would run VAE detection for all heuristics from 20180101 until the present.

### Running the VAE listing

Similarly, running VAE listing is done from the command line:

{$ESPHOM]/bin/esp vae\_listing

The command will accept one optional argument: use -p to generate a listing containing PHI, otherwise the report will obscure all PHI. For the default report, dates are provided as only years, dates of service are provided as days offset from the vaccination date. No patient information is provided.

The listing should be reviewed to confirm that the system configuration is generating VAEs correctly. The PHI version of the listing can be used to conduct chart review to examine the potential VAE cases to determine

## Section 4: MDM-T02 message interface

Once VAE cases have been generated, subsequent command can be used to generate HL7 version 2.3.1 MDM-T02 messages, which contain a brief report concerning the potential VAE. These messages are generated by ESP, but must be transferred to the EHR system and imported. The MDP-T02 message is a standard message format for providing external documents regarding a patient clinical issue, to the attention of a physician. The process for transfer of the messages must be designed by the site, with coordination between the ESP system administrator and the EHR HL7 interface developer.

Once these messages are loading into the EHR system, the notified care provider will have a brief message describing the potential VAE, and a URL link back to an ESP web page where the VAE data can be reviewed. The care provider will be able to confirm or reject the VAE. Confirmed VAEs will accept additional information from the care provider, and when saved the results will be used to generate a VAERS message for transfer to the VAERS reporting system.

When VAERS messages are sent, an additional MDM-T02 message is generated as well. This message informs the care provider that a VAERS message has been sent to the reporting system, and this is saved as part of the patients medical record.

## Section 5: Setting up PHINMS and sending VAERS messages.

Once VAEs have been confirmed, or for category 3 VAEs older than a week, VAERS reports may be generated for transfer to the VAERS messaging system.

VAERS messages are generated via the command:

[$ESPHOME]/bin/esp vaers\_hl7

This command has no options. VAERS message files are written to a directory location based on the ESP installation configuration.

PHINMS installation and operation is documented extensively elsewhere: <https://www.cdc.gov/phin/tools/phinms/installation.html>.

1. A separate project is developing an ESP data load system to use HL7 CCD documents. This capability will be ready by fall, 2019 and could be used for ESP-VAERS system installations. [↑](#footnote-ref-1)