

# User Generated SQL Queries Inform Evaluation of NSSP ESSENCE

Aaron Kite-Powell\*, Michael Coletta and Jamie Smimble

CDC, Atlanta, GA, USA

## Objective

The objective of this work is to describe the use and performance of the NSSP ESSENCE system by analyzing the structured query language (SQL) logs generated by users of the National Syndromic Surveillance Program's (NSSP) Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE).

## Introduction

As system users develop queries within ESSENCE, they step through the user-interface to select data sources and parameters needed for their query. Then they select from the available output options (e.g., time series, table builder, data details). These activities execute a SQL query on the database, the majority of which are saved in a log so that system developers can troubleshoot problems. Secondarily, these data can be used as a form of web analytics to describe user query choices, query volume, query execution time, and develop an understanding of ESSENCE query patterns.

## Methods

ESSENCE SQL query logs were extracted from April 1, 2016 to August 23, 2017. Overall query volume was assessed by summarizing volume of queries over time (e.g., by hour, day, and week), and by Site. To better understand system performance the mean, median, and maximum query execution times were summarized over time and by Site. SQL query text was parsed so that we could isolate, 1) Syndromes queried, 2) Sub-syndromes queried, 3) Keyword categories queried, and 4) Free text query terms used. Syndromes, sub-syndromes, and keyword categories were tabulated in total and by Site. Frequencies of free text query terms were analyzed using n-grams, wordclouds, and term co-occurrence relationships. Term co-occurrence network graphs were used to visualize the structure and relationships among terms.

## Results

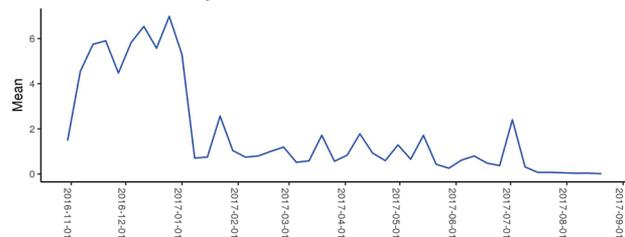
There were a total of 354,101 SQL queries generated by users of ESSENCE between April 1, 2016 and August 23<sup>rd</sup>, 2017. Over this entire time period there was a weekly mean of 4,785 SQL queries performed by users. When looking at 2017 data through August 23 this figure increases to a mean of 7,618 SQL queries per week for 2017, and since May 2017 the mean number of SQL queries has increased to 10,485 per week. The maximum number of user generated SQL queries in a week was 29,173. The mean, median, and maximum query execution times for all data was 0.61 minutes, 0 minutes, and 365 minutes, respectively. When looking at only queries with a free text component the mean query execution time increases slightly to 0.94 minutes, though the median is still 0 minutes. The peak usage period based on number of SQL queries performed is between 12:00pm and 3:00pm EST.

## Conclusions

The use of NSSP ESSENCE has grown since implementation. This is the first time the ESSENCE system has been used at a National level with this volume of data, and number of users. Our focus to date has been on successfully on-boarding new Sites so that they can benefit from use of the available tools, providing trainings to new

users, and optimizing ESSENCE performance. Routine analysis of the ESSENCE SQL logs can assist us in understanding how the system is being used, how well it is performing, and in evaluating our system optimization efforts.

NSSP-ESSENCE Weekly Mean Execution Time in Minutes



## Keywords

evaluation; system usage; public health practice

\*Aaron Kite-Powell

E-mail: lyv8@cdc.gov

