ISDS
INTERNATIONAL SOCIETY
for DISEASE SURVEILLANCE

Development of Automated Data Quality Indicators and Visualizations using Florida's ESSENCE System

Wayne Loschen*[1] and Aaron Kite-Powell[2]

[1]Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA; [2]Florida Department of Health, Jacksonville, FL, USA

Objective

The objective of this project was to develop visualizations and tools for public health users to determine the quality of their surveillance data. Users should be able to determine or be warned when significant changes have occurred to their data streams, such as a hospital converting from a free-text chief complaint to a pick list. Other data quality factors, such as individual variable completeness and consistency in how values are mapped to standard system selections should be available to users. Once built, these new visualizations should also be evaluated to determine their usefulness in a production disease surveillance system.

Introduction

Understanding your data is a fundamental pillar of disease surveillance success. With the increase in automated, electronic surveillance tools many public health users have begun to rely on those tools to produce reports that contain processed results to perform their daily jobs. These tools can focus on the algorithm or visualizations needed to produce the report, and can easily overlook the quality of the incoming data. The phrase "garbage in, garbage out" is often used to describe the value of reports when the incoming data is not of high quality. There is a need then, for systems and tools that help users determine the quality of incoming data.

Methods

A series of data quality visualizations were developed and implemented in the Florida Department of Health's version of ESSENCE. Users were given numerous pages that showed different aspects of data quality, such as variable-level percent completeness measurements shown by hospital or county. Other items included the percent of time a value that should have been a part of a specific reference list was actually present and matched to known values, the number of input files received by a hospital, and the time each data source was processed. Finally, an algorithm and visualization was developed to alert users when data quality factors had changed significantly. With

access to all these new screens, users of the system were given the opportunity to use the system and their usage and opinions were collected.

Results

The data quality portal has been active the Florida ESSENCE system since March 31st, 2012. Between that time and August the portal has been accessed over 1300 times. The presentation will include additional statistics about which specific features were most used and those features that users have found the most useful. In addition, data quality issues that have been discovered using the new tool will be discussed.

Conclusions

With the ever increasing amount of data that public health must analyze due to meaningful use, it is imperative that tools and visualizations that can decipher data quality issues be made available in an easily accessible format and without the need for tools external to the system. If systems continue to ignore changes in the data they receive automatically, it can easily produce degraded or incorrect analyses and interpretation of events, leading to wasted resources. These results can negatively impact the decisions and responses that public health users make, especially in light of the increasing reliance on these types of systems for up to the minute information. This project has developed tools and visualizations that can help determine the data quality issues as they are occurring. This presentation will outline the lessons learned for using and creating these tools so they can be shared for everyone to benefit.

Keywords

Visualization; Data Quality; Meaningful Use

*Wayne Loschen
E-mail: wayne.loschen@jhuapl.edu