ISDS 2014 Conference Abstracts

Surveillance of Surveillance: Inventorying Gaps and Commonalities Across the Universe of Surveillance Systems

Catherine Ordun* and Atlanta Health Informatics Alliance

Strategic Innovation Group, Booz Allen Hamilton, Decatur, GA, USA

Objective

To identify analytic gaps and duplication across U.S. government, international agencies, non-profit and academic health surveillance systems, programs, and initiatives in four areas: Analytics, Data Sources, Statistics, and System Requirements.

Introduction

Health surveillance systems provide important functionalities to detect, monitor, respond, prevent, and report on a variety of conditions across multiple owners. They offer important capabilities, with some of the most fundamental including data warehousing and transfer, descriptive statistics, geographic analysis, and data mining and querying. We observe that while there is significant variety among surveillance systems, many may still report duplicative data sources, use basic forms of analysis, and provide rudimentary functionality.
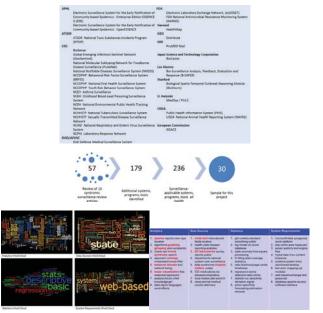
Methods

We conducted a review of 15 "review articles" revealing 57 biosurveillance systems dedicated to infectious disease and clinical surveillance. In a second review, we identified an additional 179 systems, programs, initiatives, and databases yielding 236 systems, programs, initiatives, and databases that have been labeled as "surveillance systems", across multiple U.S. government agencies and health conditions. We chose 30 out of 236 systems based on availability of public data, completeness of information about the program, and longevity of the system, for a preliminary inventory analysis. Next, we conducted topic modeling using Mallet open source software from Boston University to identify key terms collected for all 30 systems that revealed common terms across four categories: Analytics, Data Sources, Statistics, and System Requirements.

Results

Thirty systems, programs, and initiatives were identified across 14 owners, with the largest representation from the Centers for Disease Control and Prevention (CDC). Topic modeling revealed the following key terms and using Wordle.net, we visualized the relationship and frequency of the occurrence of these topics by generating word clouds, by each category. Common analytics: text, maps, temporal analysis, charts, and some form of trend analysis and algorithms. Common data sources: state-based data, use of news, clinical data and laboratory data. Common statistics: basic descriptive methods; gaps evident in more advanced analysis. Common system requirements: web-based portals; lack of specific IT architectures such as open source, or cloud-based systems.

Conclusions

The 30 systems ranged from very basic surveillance collected through surveys via local health departments, to sophisticated natural language processing algorithms to collect open source data. Our topic modeling confirmed our understanding that the majority of surveillance systems are still basic - web-based portals, general analytics, basic visualizations, some descriptive statistics. We are

exploring a more rigorous methodology to conduct a "gaps and commonalities" analysis across all 236 systems and seek to develop a ratings system that scores across each of the four categories we reviewed (Analytics, Data Sources, Statistics, and System Requirements) as an evaluation framework for future analysis.

Keywords

surveillance system; topic modeling; georgia tech; knowledge gaps; emory university

*Catherine Ordun
E-mail: ordun_catherine@bah.com