ISDS 2016 Conference Abstracts

An open source inventory to evaluate public health surveillance systems

KC Decker[1], Catherine Ordun[2] and Dimitrious G. Koutsonanos*[1, 2]

[1]Rollins School of Public Health, Emory University, Atlanta, GA, USA; [2]Booz Allen Hamilton, McLean, VA, USA

Objective

The objective of this project is to advance the science of biosurveillance by providing a user curated cataloging system, to be used across health department and other users, that advances daily surveillance operations by better characterizing three key issues in available surveillance systems: duplication in biosurveillance activities; differing perspectives and analyses of the same data; and inadequate information sharing.

Introduction

A variety of government reports have cited challenges in coordinating national biosurveillance efforts at strategic and tactical levels. The General Accountability Office (GAO), an independent nonpartisan agency that investigates how the federal government funding and performs analysis at the request of congressional committees or by public mandate, has published 64 reports on biosurveillance since 2005. The aim of this project is to better characterize these issues by collecting and analyzing a sample of publicly documented biosurveillance systems, and making our data and results available for the public health community to review and evaluate. This study openly publishes the data files of information collected (i.e. CSV, XLS), the Python NLP scripts, and a freely available web-based application developed in R Shiny that filters against the 227 biosurveillance systems and activities to promote a more transparent understanding of how public health practitioners conduct surveillance activities.

Methods

Collected and reviewed data on 424 systems, of which 227 systems and activities met our criteria; Implemented a new approach to develop a standard framework for data collection using natural language processing (NLP); Openly published all data files publicly on Github and developed an online analytics application; and Convened a workshop of experts from across federal, state, not-for-profit, academic and commercial entities in November 2015 in Washington, D.C., to review the methodology and results of this study.

Results

The results of this project include a fully functional web application and code (available through Github) for the continued expansion, categorization and analysis of surveillance systems. Unique findings currently rendered through the 227 surveillance systems include: Out of 227 systems, 20 were established in the year 2006, alone, with an increase in systems established following 1990; 68% of all systems catalogued are focused solely on human surveillance; 45% of all cataloged systems used statistical analysis and only 4% are using Natural Language Processing; and 43% of all biosurveillance systems in our inventory reported using "health department" data as a data source.

Conclusions

We believe this project is the first step for public health practitioners and researchers to contribute to a transparent inventory of systems and activities. Results provide meaningful metadata on an over focus on human surveillance, over-reliance on a single data source (health departments) and a lack of advanced data science practices being applied to systems in the field. The value of this project 1) provides a starting point for the development of a standard framework of categories to use for cataloging biosurveillance systems, 2) offers openly available data and code on Github [3] for others to integrate into their research, and 3) introduces a set of methodological issues to consider in a biosurveillance inventorying exercise.

Top ten data dissemination groups by number of systems

Data dissemination groups	Number of biosurveillance systems	% Total
Online	92	40.9%
Reports	72	32.0%
Public Access	29	12.9%
Annual Reports	29	12.9%
Periodicals	22	9.8%
Files	19	8.4%
Alerts	17	7.6%
Publications	17	7.6%
Summaries	16	7.1%
Database	15	6.7%

Top ten Conditions groups by number of systems and activities

Condition groups	Number of biosurveillance systems	19.6%
Influenza	44	19.6%
Infections	36	16.0%
Animal	26	11.6%
Virus	25	11.1%
Injury	23	10.2%
Fever	22	9.8%
Respiratory Disease	20	8.9%
Mortality	18	8.0%
Illness	18	8.0%
Syndromes	18	8.0%

Keywords

Biosurveillance; Data Science; Open Source; Analytics; WebApplication

Acknowledgments

References

1. U.S. Government Accountability Office. GAO-15-664T, Biosurveillance: Additional Planning, Oversight, and Coordination Needed to Enhance National Capability [Internet]. Washington (DC): U.S. Government Accountability Office; 2015 [Cited 2016Mar02]. Retrieved from: http://gao.gov/products/gao-15-664t

2. National Biodefense Science Board. Recommendations for Action: Modernizing and Enhancing Our Nation's Biosurveillance Capabilities Report from the National Biodefence Science Board [Document on the Internet]. Washington (DC): U.S. Department of Health and Human Services, Assistant Secretary for Preparedness and Response; 2015. [Cited 2016Mar02]. Retrieved from: http://www.phe.gov/Preparedness/legal/boards/nprsb/recommendations/Documents/biosrveillance-capabilities.pdf

3. Phillips N. Development of Python scripts to classify surveillance systems; 2015 [Repository on the Internet]. Retrieved from: https://github.com/nick-phillips/SoS, web application at http://bit.ly/1HuNkET

***Dimitrious G. Koutsonanos**
E-mail: dimkoy@gmail.com