ISDS 2019 Conference Abstracts

Data Science, Analytics and Collaboration for a Biosurveillance Ecosystem

Karen A. Stark, Amol Shah, Jacob Borgman, Miko Somborac, Jeremy Carson, Loren Hauser, Krishna Kola, Hemant Virkar

Digital Infuzion, Gaithersburg, Maryland, United States

Objective

While there is a growing torrent of data that disease surveillance could leverage, few effective tools exist to help public health professionals make sense of this data or that provide secure work-sharing and communication. Meanwhile, our ever more-connected world provides an increasingly receptive environment for diseases to emerge and spread rapidly making early warning and collaborative decision-making essential to saving lives and reducing the impact of outbreaks. Digital Infuzion's previous work on the Defense Threat Reduction Agency (DTRA)'s Biosurveillance Ecosystem (BSVE) built a cloud-based platform to ingest big data with analytics to provide users a robust surveillance environment. We next enhanced the BSVE data sources and analytics to support an integrated One Health paradigm. The resulting BSVE and Digital Infuzion's HARBINGER platform include: 1) identifying and ingesting data sources that span global human, animal and crop health; 2) inclusion of non-health data such as travel, weather, and infrastructure; 3) the data science tools, analytics and visualizations to make these data useful and 4) a fully-featured Collaboration Center for secure work-sharing and communication across agencies.

Introduction

After the 2009 H1N1 pandemic, the Assistant Secretary of Defense for Nuclear, Chemical and Biological Defense indicated “biodefense” would include emerging infectious disease. In response, DTRA launched an initiative for an innovative, rapidly emerging capability to enable real-time biosurveillance for early warning and course of action analysis. Through competitive prototyping, DTRA selected Digital Infuzion to develop the platform and next generation analytics. This work was extended to enhance collaboration capabilities and to harness data science and advanced analytics for multi-disciplinary surveillance including climate, crop, and animal as well as human data. New analysis tools ensure the BSVE supports a One Health paradigm to best inform public health action. Digital Infuzion and DTRA first introduced the BSVE to the ISDS community at the 2013 annual conference SWAP Meet. Digital Infuzion is pleased to present the mature platform to this community again as it is now a fully developed capability undergoing FedRAMP certification with the Department of Homeland Security’s National Biosurveillance Integration Center and Is the basis for Digital Infuzion's HARBINGER ecosystem for biosurveillance.

Methods

We integrated over 170 global One Health data sources using cloud-based automated data ingestion workflows that provide unified access with data provenance. We used modular automated workflows to implement data science including Natural Language Processing (NLP), machine learning, anomaly detection, and expert systems for extraction of concepts from unstructured text. A first of its kind ontology for biosurveillance permits linking of data across sources. This ontology allows users to rapidly find all relevant data by looking at semantic relationships within and across data sets having varying quality, types, and usages to understand the best, most complete indicators of impending threats.

We applied the following principles to the development of data science tools: 1) mathematics should be fully automated and operate 'under the hood' without need for user intervention; 2) 'At-a-Glance' visualizations should summarize Information, draw attention to key aspects and permit drill down into underlying data; 3) data science analytics and tools need to be validated with real-world data and by disease surveillance experts and 4) secure collaboration capabilities are essential to biosurveillance activities. This was a highly complex effort. We worked closely with surveillance analysts from multiple agencies and organizations to continuously guide the development of capabilities. We drew upon subject matter expertise in public health, machine learning, social media, NLP, semantics, big data integration, computational science, and visualization. A high level of automation, security and immediacy of data was applied to support rapid identification and investigation of potential outbreaks.

Results

The platform now provisions integrated One Health information. Data sources were harmonized and expanded, along with historical information, to better predict and understand biothreats. These include global social media, human, plant, animal, and weather data. An Analyst Workbench delivers logical, intuitive and interactive visualizations enabling disease surveillance professionals to identify critical, predictive information without extensive manual research. Over 700 approved users currently have access to the prototype. Biosurveillance activities can be performed collaboratively among governmental agencies, public health officials, and the general public using the Collaboration Center and its sharing and messaging systems. Data sharing is HIPAA compliant and distinguishes public from private data using carefully controlled and approved role- and attribute-based access for security. To speed disease surveillance workflows, the workbench generates suggestions to the user on their current work. Anomaly detection to alert to potential developing disease events employs fully automated analytics to conduct over 43 million calculations daily for more than 500 diseases in over 170 data sources, distilling this into a table that ranks the most significant anomalous increases that may indicate an outbreak and warrant investigation. A predictive disease modeling tool based on current and historical data uses fuzzy logic to identify the likeliest outcome, even early in an outbreak when there is much uncertainty about the disease and its characteristics. A complex automated workflow identifies health-related topics that are trending in Twitter and evaluates their severity using novel lexicons and new reactive sentiment analysis. Searches use the ontology to gather all relevant information and are supported by the most advanced NLP with custom surveillance rules to provide succinctly extracted information. This alleviates the need for extensive reading by identifying exactly which data is needed and extracting key concepts from it. Intuitive methods of visual representation, interactive displays, and drill-down capabilities were leveraged in all analytics for rapid understanding of results. Finally, we added a software development kit to enable third party developers to continuously enhance the platform capabilities by adding new data sources and new analytic apps. This allows the platform to be adapted for specific needs and to keep pace with new scientific and technical discoveries and has resulted in over 50 analytic apps.

Conclusions

The addition of One Health data and analytics, and the integration of health data with unconventional data sources and modern approaches to data science and complex workflows, resulted in enhanced situational awareness and decision-making capabilities for users. The expanded Collaboration Center within the workbench, enables users to partner and collaborate with other agencies and biosurveillance professionals both nationally and internationally to maximize the rapidity of responses to serious disease outbreaks.

Acknowledgement

This project was supported by the Defense Threat Reduction Agency (DTRA) and the Department of Homeland Security (DHS) National Biosurveillance Integration Center (NBIC) via contracts to Digital Infuzion, Inc.