Affordable System for Rapid Detection and Mitigation of Emerging Diseases

Article *in* International Journal of E-Health and Medical Communications · January 2011

DOI: 10.4018/jehmc.2011010105 · Source: DBLP

Affordable System for Rapid Detection and Mitigation of Emerging Diseases

Nuwan Waidyanatha, LIRNEasia, China

Artur Dubrawski, Carnegie Mellon University, USA

Ganesan M., IITM's Rural Technology and Business Incubator, India

Gordon Gow, University of Alberta, Canada

ABSTRACT

South and South-East Asian countries are currently in the midst of a new epidemic of Dengue Fever. This paper presents disease surveillance systems in Sri Lanka and India, monitoring a handful of communicable diseases termed as notifiable. These systems typically require 15-30 days to communicate field data to the central Epidemiology Units, to be then manually processed (Prashant & Waidyanatha, 2010). Currently used analyses rely on aggregating counts of notifiable disease cases by district, disease, and week. The Real-Time Biosurveillance Program (RTBP), a multi-partner initiative, aims at addressing those challenges by developing affordable paradigm-changing Information Communication Technology (ICT), implementing and field-testing them in India and Sri Lanka. Key components of the proposed solution include real-time digitization of clinical information at hospitals and clinics with the mHealthSurvey mobile phone software (Kannan et al., 2010), detecting anomalies in large multivariate biosurveillance data using the T-Cube Web Interface spatio-temporal statistical analysis tool (Ray et al., 2008), and disseminating critical information pertaining to the adverse events to healthcare workers using the Sahana Alerting Module (Sampath et al., 2010). This paper provides an overview of the applications and discusses utility of the technologies for real-time detection and mitigation of emerging threats to public health.

Keywords: *Alerting, Biosurveillance, Data Mining, Epidemiology, Interoperability, Mhealth, Spatio-Temporal Analysis*

INTRODUCTION

Public health monitoring systems used currently in India and Sri Lanka focus on a subset of approximately 25 communicable diseases designated as notifiable. When a patient is diagnosed with one of those diseases, a medical officer notifies regional health administration or local authorities using paper-based reporting methods and follows up with a phone call in certain instances. Health Inspectors would then make house calls to investigate the individual

DOI: 10.4018/jehmc.2011010105

cases and, if needed, engage in mitigation and prevention procedures.

Epidemiological statistics streaming from hospitals, health clinics, and health centers are gathered on paper forms and mailed in hard-copies. Upon receipt by the regional health offices, the data is periodically consolidated by departmental staff and summaries are passed further up the organizational structures, still on paper. Then, regional and national epidemiological centers can use the aggregated data for trending to identify any long term effects that might require reallocation of resources.

In certain instances of the current process computers are used sparingly and primarily for generation of printed reports. Therefore, potential efficiencies due to automation or digitization are not realized. The Indian National Informatics Center (NIC) hosted Directorate of Public Health and Preventive Medicine (DPH&PM) web portal is a tool for entering manually summarized weekly or monthly data aggregates for generating electronic "Morbidity/Mortality" reports. The Sri Lankan Epidemiology Unit manually enters aggregated weekly counts received from health departments into spreadsheets to produce the "Weekly Epidemiological Returns" electronic report. Under the present system, it can take up to 30 days for information to move through, yielding avoidable delays in outbreak detection, notification, and response.

Limited efficiency of the present systems is therefore an identified and important challenge. Our observations and interviews with health officials (Dr. P. Hemachandra - Regional Epidemiologist and Dr. L. Edirisinghe - Deputy Director Planning, Ministry of Health and Nutrition, Sri Lanka, consulted 16 December 2009; Dr. N. Raghupathy – Deputy Director of Health Services, Tamil Nadu Department of Health and Family Welfare, India) reveal that only about 20% of the patient preliminary diagnoses are confirmed by a medical officer, while the remaining 80% are left in probable or suspected status. Hospitals and clinics in rural India and Sri Lanka that examine a hundred patients each day operate predominantly as if they were clearing houses. Patient arrivals dominate the morning hours, often forcing reduction of service times to less than 2 minutes per person and posing additional challenges to accuracy of available and reported epidemiological information.

Leading experts in the field of biosurveillance and health informatics have argued that improvements in disease detection and notification can be achieved by introducing more efficient means of gathering, analyzing, and reporting of data (Wagner, 2006). Introduction of new Information and Communication Technologies (ICT) create opportunities to improve efficiency in an affordable fashion, and affordability is the key requirement for successful deployments of any technology in developing countries. The primary research objective of our multi-partner research initiative, Real-Time Biosurveillance Program (RTBP), is to examine those opportunities more closely through a field-test-driven research into integration of the new ICTs with existing disease surveillance and detection processes (Vital Wave Consulting – report, 2009; Neill & Moore, 2006; Chapman, 2009).

We have established that it is possible to use RTBP approach to substantially shorten the latency of disease reporting, to make disease surveillance much more comprehensive and expressive than before, and to improve situation awareness of public health decision makers through the use of the alerting subsystem. In the process, we had to overcome numerous challenges, and the iterative evolution of the system continues. The attainable benefits of merit to the society include new efficiencies in public health event detection, reduced costs in gathering and processing health statistics, increased scope and depth of these statistics, enabled optimization of allocation of medical resources, and the ability to improve planning of short and long term preventive actions (Gow et al., 2010).

TECHNOLOGIES

Digitizing Patient Data with the mHealthSurvey

Initial and an important step of the RTBP process are collecting reliable patient data from hospitals and clinics in order to conduct epidemiological analyses. Mobile phones present an opportunity to both digitize and transmit patient data using existing commercial wireless data service in a relatively cost-effective manner. The use of commercial wireless services to support health care initiatives in developing countries is gaining recognition within an emerging field known as “mHealth” (Vital Wave Consulting, 2009).

The mHealthSurvey was designed to operate on standard affordable mobile phones. It has been developed through the RTBP project by a team at the Indian Institute of Technology Madras’s Rural Technology and Business Incubator as a means of collecting real-time patient disease, syndrome, and demographic data. The decision to develop the mHealthSurvey, instead borrowing one of the available disease specific mHealth solutions, was based on the need to assure interoperability with other components of the biosurveillance system and to install and operate it on any low cost Java- enabled mobile phone.

The RTBP mHealthSurvey has three basic steps, which are installing and customizing the software for the healthcare worker by creating a simple user profile; digitizing the patient epidemiological data, and delivering the patient records over the GSM Internet channels. Patient data, transmitted over the mobile operator’s network to an Internet gateway, goes to a central database for storage and analysis. The engineering components, architecture, and functionality of the mHealthSurvey mobile application is discussed by (Kannan et al., 2010; Gow et al., 2010).

- The mHealthSurvey application is available for users to download from a specific location in the cyberspace, which is typically done by entering the URL into the phone’s Wireless Application Protocol (WAP) browser. The browser retrieves the Java Archive (JAR) file through the General Packet Radio Service (GPRS). Once downloaded the phone will prompt the user the installed the software on the phone.
- After installing the application, the users execute the download list function which retrieves the lookup values from a database that includes pre-defined lists of disease names, signs, symptoms, age-groups, gender names, case-status, location types, and health worker types (Figure 1 (a)). This is usually a one-time step but the download feature should be executed periodically to update the list of diseases, signs, and symptoms in the mobile phone memory.
- The next step is a profile registration process (Figure 1(b)) to generate a universal unique identifier for the user from the database to the mHealthSurvey application. The purpose of the identifier is to tag all records received from the unique user. The application allows a single phone to register more than one health worker profile so that several healthcare workers at a health facility can share a single mobile phone.
- User must identify the geographic location at which they work. The health worker selects their jurisdiction type; i.e., the administrative geographic area name; followed by entering the name of that area (Figure 1(c)). For example, location type = “PHC” and the location name =“Sevanipatti” would retrieve all the locations belonging to the Sevanipatti Primary Health division.

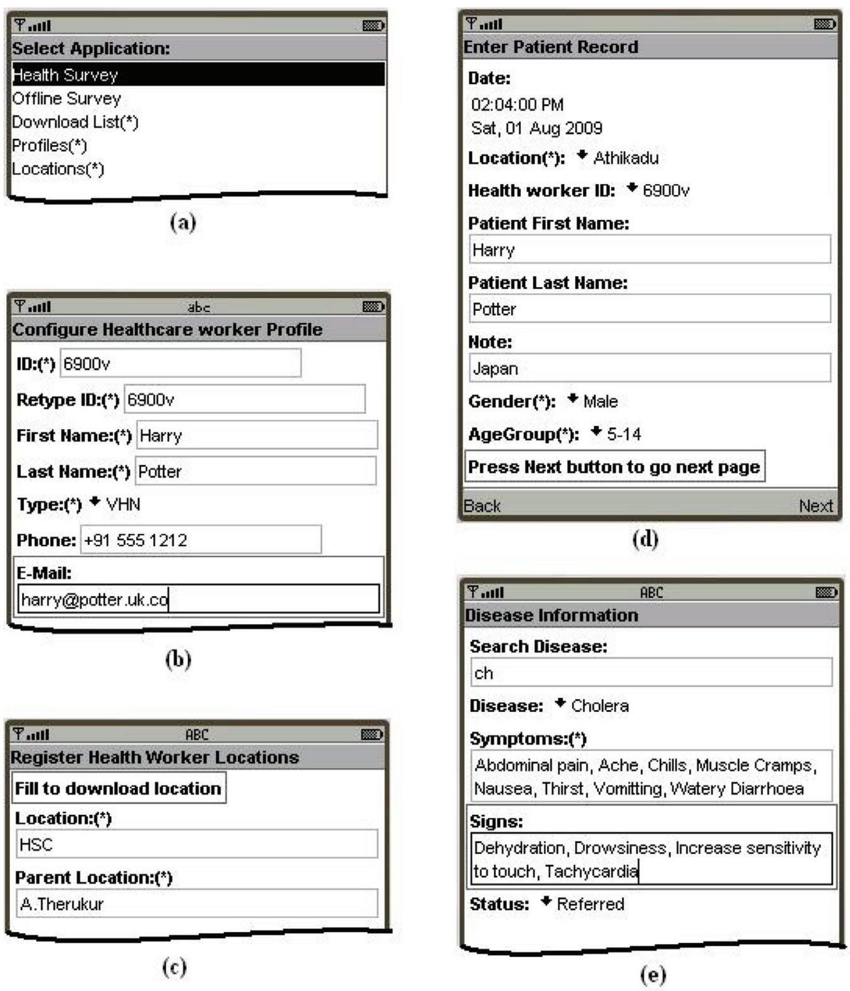
After proceeding through steps 1 – 4 of installation and configuration, the application presents the healthcare worker with a series of fields and menus that are completed using the standard phone keypad. Patient data is captured with several attributes included in the health survey form: case date and time; health worker id; location name; patient first name; patient last name; notes; gender; age-group; disease; symptoms; signs; and case-status (Figure 1(d) and 1(e)).

Case date and time will default to the current date and time, as set in the mobile phone, anticipating that health workers are sending data in almost real-time. However, the application does give users the option of overriding this feature and inserting customized date and time, should there be a delay in submitting the records. Health worker id, location name, age-group, gender, and status can be selected from pre-defined drop-down lists that are initially set through the download list function explained as Step 2 in the installation and configuration process.

The application includes several design features intended to simplify the data entry. For example, pre-defined menus of disease, symptom, and sign types are incorporated, so the users need only to begin entering the first few characters of the required value in the appropriate field to bring up a list of matching names (Figure 1(e)).

The mHealthSurvey client application communicates data to the central database

Figure 1. mHealthSurvey Screens (a) main menu (b) profile (c) location (d) patient record page 1 (e) patient record page 2



through Hyper Text Transfer Protocol (HTTP) Post, Request, and Get functions embedded on the server side PHP hypertext pre-processor application that follows a REST (Representation State Transfer) like architecture. Each record is typically about 2kb in size. The transmission cost per month per participating health facility can be kept under 2.00 American dollars in both Sri Lanka and India.

T-Cube Web Interface for Anomaly Detection

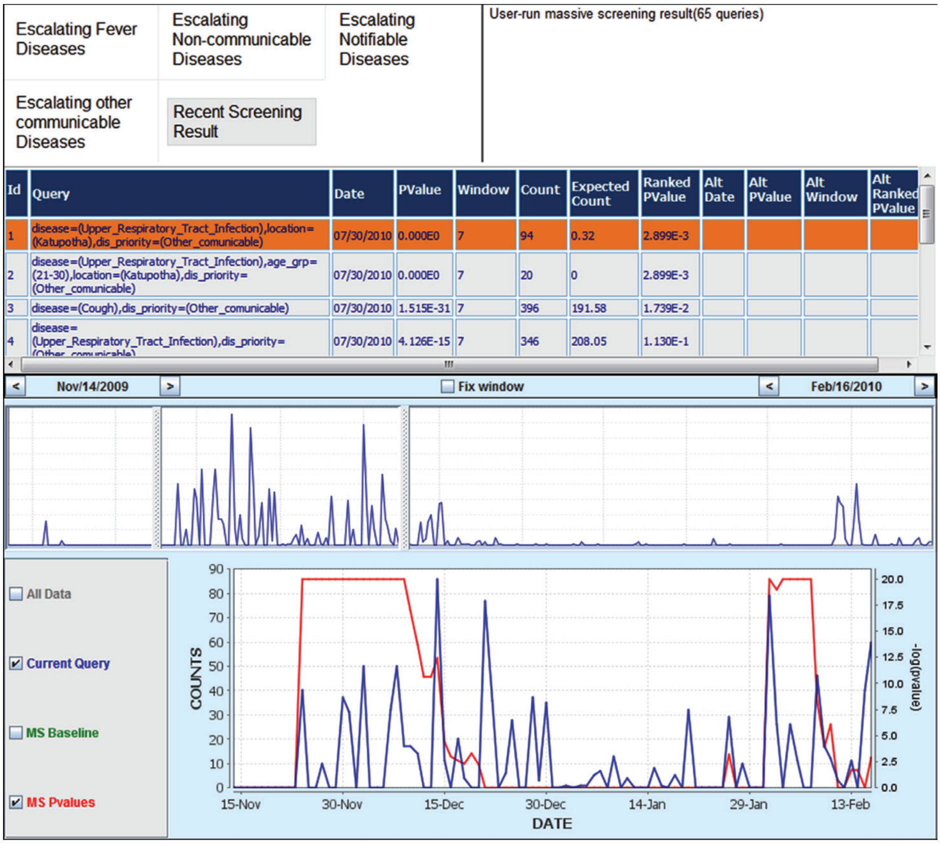
T-Cube Web Interface (TCWI), engineered at Carnegie Mellon University Auton Lab, is a generic tool designed to efficiently visualize and manipulate large scale multivariate temporal and spatio-temporal datasets commonly encountered in public health applications (Ray et al., 2008). The interface allows the user to execute complex ad-hoc queries quickly and to run various types of statistical tests on the loaded data. Upon uploading the working dataset, the user can manipulate and visualize data through the Time Series, Map, and Pivot Table panels.

The user may choose to apply one of the available statistical modeling and anomaly detection techniques. The list of choices includes moving average, moving sum, cumulative-sum, temporal scan, change scan, linear trend, peak analysis and range analysis. The users can interactively manipulate, navigate, summarize and visualize data at interactive speeds. That supports focused investigations, drill-downs as well as summarizing and reporting operations. The users may choose to simply execute a Massive Screening procedure, which performs an automatic and comprehensive search for anomalous patterns across large number of queries spanning multiple dimensions of data. This function could be invoked interactively by the user, or it could be scheduled to execute periodically to generate a set of alerts. The alerts are sorted according to statistical significance of the corresponding anomalies found in data, and they can be interactively reviewed by the Epidemiologists for the factual confirmation of their practical importance. TCWI supports these efforts by allowing focusing attention on the most surprising patterns in the current data, and by providing the ability to quickly drill down or roll up the data for further explanations. Figure 2 shows a list of the statistically significant events with the time series for one (highlighted) alert indicative of a potential Respiratory Tract Infection outbreak in Kurunegala District of Sri Lanka.

The core of the TCWI Massive Screening procedure is the Temporal Scan bi-variate anomaly detection algorithm (Dubrawski, 2010). It leverages the efficiency of the T-Cube data structure to perform a massive number of tests of statistical hypotheses in order to find the most significantly anomalous patterns in data. The TCWI also implements Multivariate Bayesian Spatial Scan algorithm (Neill & Cooper, 2010) that complements the analyzes by testing the spatial as well as temporal correlations between health events. This algorithm computes the overall probability of a disease outbreak anywhere in the scope of data selected by the user, separately for each day within that scope. The national score for the current day is reported in the upper right corner of the map display window, shown in Figure 3.

Figure 3 shows an example detection of a potential outbreak of food poisoning. The spatial and temporal distributions of the corresponding recorded disease cases are shown in blue. The history of the estimated probability of the food poisoning occurring on a given day anywhere in the nation is depicted with the red line plot. Massive screening automatically identifies the periods of time of abnormally high frequency of cases of food poisoning, relative to cases of other diseases reported nationally. Spatial distribution of probabilities of the food poisoning outbreak computed for separate regions of Sri Lanka for the current day is depicted with red circles sized according to the value of the estimated probability. In this example, a central East-West swath of the country seems to be primarily affected by this disease. Figure 3 shows a global score of 0.9795 (i.e. almost a 98% chance) for food poisoning event on August 4th, 2008. This event was later

Figure 2. Pre-Screening results for non-notifiable communicable diseases. The selected alert corresponds to a potential outbreak of the respiratory tract infections in Kurunegala District of Sri Lanka.



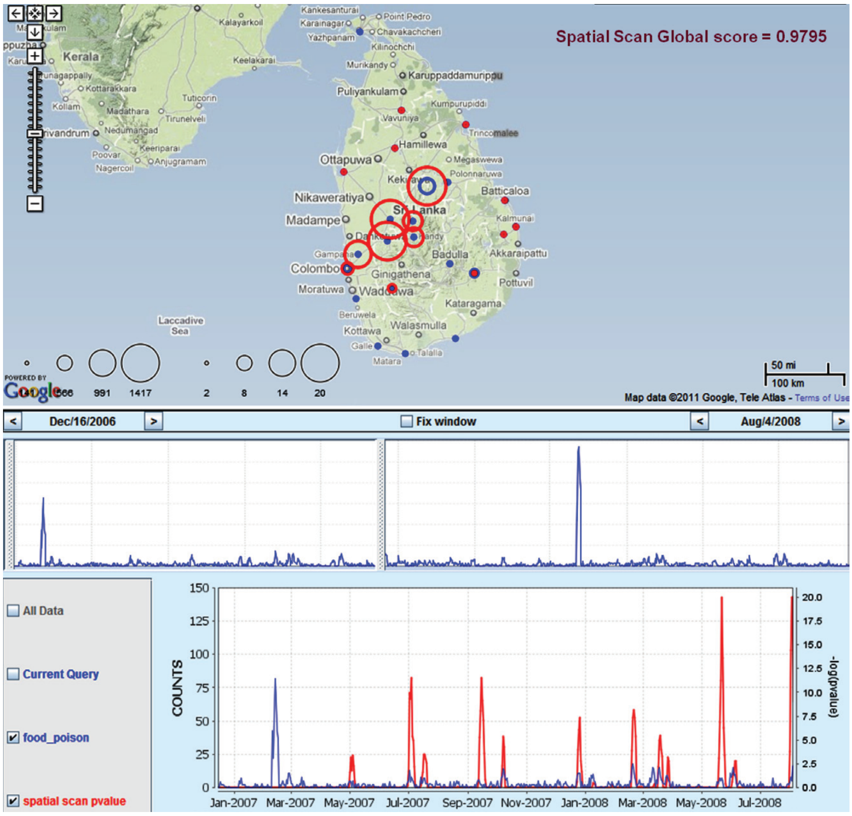
confirmed as an actual food poisoning outbreak in Kurunegala District. When prompted by the system to the new discoveries, the analysts can further drill into the data by narrowing their filters or selecting various additional modeling techniques to confirm the outbreak.

TCWI also provides a computationally efficient, interactive data summarization capability. Multidimensional data of counts of events (such as the numbers of reported disease cases) can be aggregated into a multi-way matrix view - a pivot table. Multiple attributes can be selected to denote rows and columns of the table by dragging the corresponding attribute names from the attributes list. Once a table is created and automatically filled with values, the user can click on a cell to view the corresponding time series graph, or a pie chart depicting the frequency distribution of the underlying data.

Sahana Alerting Module for Downstream Messaging

The Alerting Module is a Sahana Module (Careem et al., 2006), developed through the RTBP by a team at Respere Lanka. At present, the module allows for the generic sending and posting of messages in the form of SMS, Email, and Web. This messaging platform adopts global standard for interoperability called the Common Alerting Protocol. Future work include adopting

Figure 3. T-cube spatio-temporal visualization map interface



other standards such Emergency Data Exchange Language (Chapman, 2009).

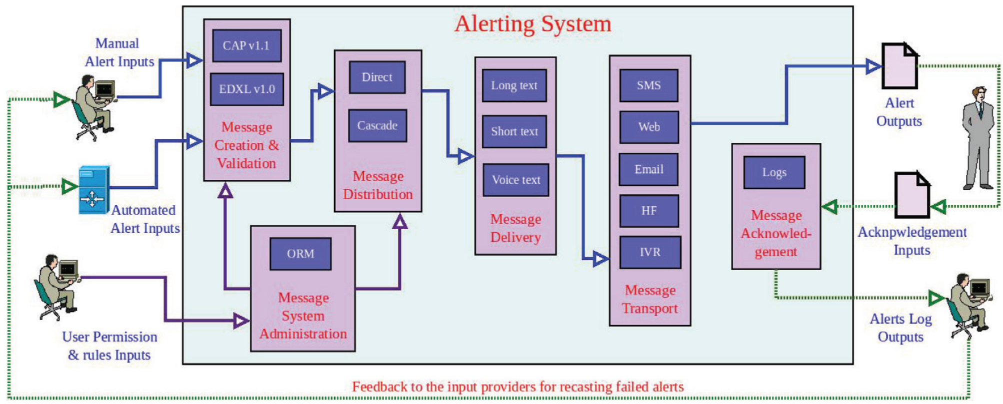
Sahana Alerting Module (SAM) is a server application that provides an intermediary point of interconnection between the health departments and health workers. The relay network facilitates interconnection of mobile phone and personal computer as well as a passage of CAP-compliant messages through a single software application. This intermediary program translates a message from the formal messaging protocol of the sender to the formal messaging protocol of the receiver in a telecommunication network where programs communicate by exchanging formally-defined messages.

The CAP compliant messaging module consist of five interconnected subsystems: Message creation and validation, Message distribution, Message delivery, Message acknowledgement, Message system administration (Figure 4).

Message creation can be done manually or it can be automated through middleware linked to the detection/decision component. A hybrid system (manual/automated) could be based on a combination of centralized and distributed architecture for message creation. An important variation on the manual/automated hybrid is the concept of cascade alerting. Once an alert or notification has been created it must then be distributed to designated recipients.

Direct alerting is the normal process in which an alerting system delivers an alert to a human recipient. Cascade alerting is a process in which an alert is sent as a system-to-system message from one jurisdiction to

Figure 4. Sahana alerting module components



another; the receiving system then distributes the alert to the appropriate recipients within the receiving jurisdiction.

Messages must reach their destination through an appropriate receiving device, be it a mobile phone, desktop PC, or other means. The contents of the CAP XML message is rendered into human readable form while taking into account the limits of bandwidth, processor capabilities, and message display constraints inherent in any particular device. Message delivery options are categorized into three types: Long text—content rendered in a form appropriate for email, fax, or web presentation; Short text—content rendered in a form appropriate for SMS, MMS, and other limited text length technologies; Voice text—content rendered in a form appropriate for voice delivery or automated voice delivery by telephone.

In some cases it may be advisable or desirable to include a backchannel for message acknowledgement from recipients. This would require a communication link back to the message delivery system to collect acknowledgement receipts and present these as a report.

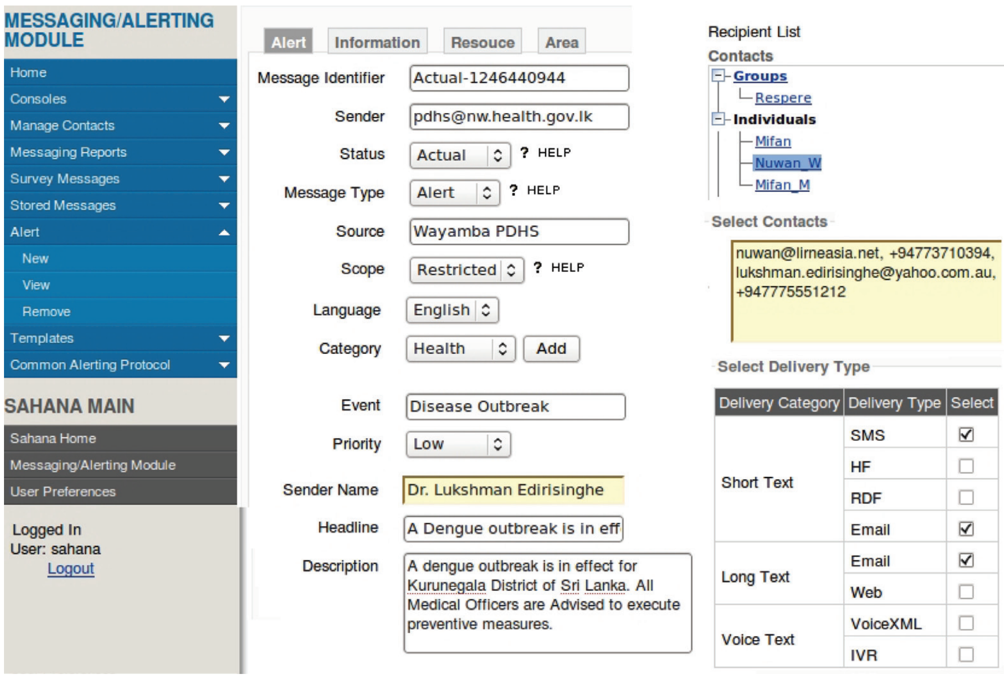
Message creation subsystem also takes into account security provisions such as user access control and originator rights management. This will require a database of user names, passwords, and ORM - Originator Rights Management (set of privileges and rules) profiles linked to the user interface software.

Figure 5 shows an example of an alert generated to notify a dengue outbreak. Each message carries a unique identifier, a set of attributes that identifies the source and sender for audits. The scope is set as restricted meaning the message is for those targeted recipients only. Category is naturally set to Health with the event described as an "outbreak". Priority defines the response actions that should be taken by the receiving health workers or health officials. If the priority was set to "urgent" then recipients may be required to take prompt action; while a "low" priority may mean being vigilant and observe the situation. The description section contains a full synopsis of the alert. Once the attributes are populated the sender can select or type in the list of recipients in the Contacts section (upper right corner in Figure 5). Thereafter, select the delivery types (or transport methods), i.e., email, SMS, web, for the message to be disseminated to the prescribed recipients via those channels.

RESEARCH DESIGN

A group of health workers, 28 in India and 15 in Sri Lanka, were supplied with a mobile phone and the mHealthSurvey application in order to digitize patient records and feed data into the RTBP. Data in India come from four Primary Health Centers and 24 Health Sub-

Figure 5. Browser interface of the Sahana alert generation screens



Centers located in the state of Tamil Nadu. In Sri Lanka, data is currently collected from a total of twelve hospitals and clinics from Kurunegala District. Health workers in India began submitting case data in June 2009. Due to delays in receiving government clearance, the Sri Lankan participants began submitting data in September 2009.

In Sri Lanka, the data is extracted from treatment chits, which are essentially pieces of paper given to each patient upon examination by a senior medical officer at the hospital or clinic. Each chit carries the name, gender, and age of the patient as well as the diagnosis (disease, symptoms, and signs), and the treatment. Research assistants belonging to a community-based primary Health Care services providing organization, Sarvodaya Suwadana Centers, visited each hospital or clinic to obtain the chits and then digitize the health records by entering the data using the mHealthSurvey application. The role of the data entry assistants was equivalent that of introducing a new human resource to the clinic or hospital, accompanying deployment of the RTBP.

The dataset in India was extracted from the outpatient registry, where the senior medical officer also documents each patient's identification number, gender, age, and diagnosis in the Outpatient Department Registry. However, in this case almost all of the healthcare workers submitting data (27 of the 29) were already working in the health centers. The senior medical officers associated with the health centers or hospitals in either country were not involved in digitizing the health records. The data digitized by the health workers is transported over the GPRS networks to a central database at a data center.

An interface between the TCWI analytic application and the database allows for health departments to access the real-time data over the World Wide Web, through a standard web browser. TCWI uses sophisticated statistical data mining algorithms to process the accumulating data each day and caches the results in the T-Cube data structure. The health department staff at the district and divisional levels access the data through the TCWI, which provides means for querying the time-series and spatial data in all shapes and forms they wish. For the ease of the health officials, a set of pre-screenings for cliques of disease, namely notifiable (reported communicable diseases), other-communicable, non-communicable, and fever-like diseases, were made available at a single push of a button (Figure 2).

The Integrated Disease Surveillance Program (IDSP) of the Deputy Director of Health Services (DDHS) in Sivaganga District as well as the Chief Nurses and Health Inspectors belonging to four Primary Health Centers, in Tamil Nadu, India, took on the responsibility of using TCWI. Similarly in Sri Lanka, the Regional Epidemiology Unit and four Medical Officer of Health Departments in Kurunegala District were trained in using TCWI. These health department staff would execute these functions each day to receive a list of potential disease outbreaks within each disease clique. They would further scrutinize this data over spatial maps to identify spread patterns, if any.

Information about the confirmed outbreaks is then disseminated to health workers and health officials in the affected and surrounding areas. This activity is done by the same health officials who detect the events through the TCWI. The SAM was implemented for this purpose, whereby messages are distributed via SMS, Email, and the Web. Typically, the messages would be received by medical officers, nurses, and health inspectors in the respective jurisdictions. There are two types of messages: 1) Action Alerts that require the healthcare workers receiving the message to execute their established response plans such as investigating the patients at their homes, educating the communities of implementing preventative measures, and 2) Situational-awareness messages which do not require immediate action but would inform the health workers to be vigilant of the potential outbreak in a neighboring area and be ready to respond if the disease were to bleed into their jurisdiction.

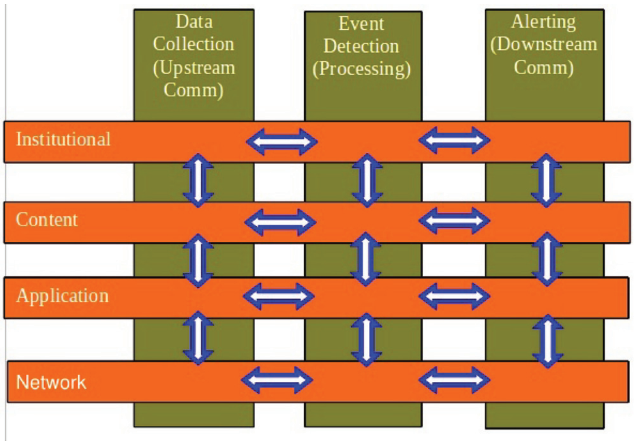
RTBP Evaluation Framework

The importance of field testing cannot be exaggerated: "Many important characteristics of biosurveillance systems can only be determined once they are deployed, at least partially, in the field" (Wagner, 2006, p. 507). Field testing can be intended as formative or summative in approach, with formative studies aimed at producing insight as to the effectiveness of a component or subsystem and identifying potential improvements to that component. The technology design of the RTBP is an integral of data collection, event detection, and alerting subsystems. The evaluation framework for field testing of biosurveillance systems examined the attributes of the system or system components (Figure 6), which can be divided into four general categories: Institutional challenges (e.g., healthcare workers, health officials, epidemiologists), Content standards (e.g., ontologies, semantics, syntax, vocabulary), Computing resources (e.g., mobile applications, detection analytics software, databases), and Communications Networks (e.g., mobile devices, computer, wireless links, Internet connections, GSM technology).

Communications networks and computing resources are evaluated based on attributes associated with technical performance. Evaluation of the institutional element of the system considered the relationships between its social and technical attributes.

To evaluate utility of the event detection functionality of the RTBP we adopted a set of generic evaluation methods known from literature (Ammenwerth et al., 2004). We have summarized a broad set of evaluation criteria on the usability of the TCWI technology, its effects on structural or process quality, problems associated with daily operation, and institutional consequences of introduction the technology. (Wagner, 2008; Lewis, 2003) have proposed biosurveillance system evaluation methods. (Anderson & Aydin, 2005) describe methods

Figure 6. Vertical components of the RTBP communication structure and horizontal layers of each component with arrows depicting the interoperability between the components and layers



and key aspects of qualitatively evaluating the organizational impact of introducing information and communication technologies in healthcare. Methods for evaluating bioinformatics systems include subjective and objective, as well as quantitative and qualitative approaches (Friedman & Wyatt, 2006).

Robustness of Digitization of Patient Data

(Gow et al., 2010) stress on acceptability being a defining attribute for the human element and explains the repercussions of usage practices among healthcare workers and the downstream consequences for data analysis and disease detection components in the RTBP. Gow et al. (2010) also emphasizes portability, defined as the likely amount of effort that will be required to introduce the system into a new location as this would be disruptive and impact on the existing procedures and practices.

We measured the robustness of the mHealthSurvey on the grounds of the quality of the data and the timeliness of delivering the digitized data. Data quality was defined as the accuracy and completeness of patient information as entered by the healthcare worker using the mHealthSurvey. This was measured in terms of the signal to noise ratio; where the signal was the entire set of data and the noise would be those records that were distorted compared to the acceptability of the record for processing. The noisy data (or erroneous) data are those that did not match the accepted spelling or the lookup values in the database. Time latency with respect to data entry was defined as the delay in which the data was submitted relative to the day of the patient's visit. Given the focus of this project as a real-time system, latency is a significant concern. Latency is measured by comparing timestamps.

Reliability of Event Detection

To evaluate the TCWI detection algorithms on timeliness and sensitivity, first, it was put to a test with past epidemiological data. This replication study incorporated, approximately, two years worth of historical epidemiological data. TCWI algorithms such as temporal scan and spatial scan have then been executed to detect unusual events in this data set. These detections were then compared against historical record of actually noticed outbreaks as reported by the epidemiological departments.

On-going comparison examines the data collected through the RTBP for possible outbreaks. These outbreaks are later verified by the epidemiologist. Lists of false alarms and true alarms are recorded to evaluate the reliability of the mHealthSurvey received data as well as the TCWI generated alerts.

Effectiveness of CAP Messaging

The CAP alerting (i.e., downstream information exchange) was a threefold evaluation process: 1) Face-to-face interviews with health officials to study the policy and procedures of the presently practiced notification system to determine the information design within the context of the CAP framework, 2) Assessment of the competency of the users knowledge of CAP and the usability of the SAM software, and 3) Evaluation of the comprehensibility of the messages by healthcare workers.

A series of interviews were conducted with healthcare workers and health officials using a set of guideline questions. These questions were designed to understand the message composition and delivery by asking them about the purpose of receiving alerts, which department should issue them, who were the decision makers, who were the message creators, what were the designations (or job titles) of the recipients, what is the message receiving organizational structure, and what were the essential content elements.

CAP issuer exercise required the participants to first interpret the event, create a CAP message, then to issue the message to the respective targeted recipients, and to give feedback on the introduced alerting component.

The message comprehension exercise comprised of the participants receiving four different messages. Each message required specific actions or inactions. The recipients of the message were asked to extract information from the messages and answer questions such as identifying the sender for authenticity, affected areas, diseases, message identifier for future reference, and priority of the message to determine the action.

DISCUSSION

Observations Regarding Data Digitization Process

The clinical data in Sri Lanka were extracted from "treatment chits". The medical officers were requested to write the diagnosis in addition to the treatment they would prescribe to the patients. Writing the diagnosis was a new component of their process, which many would forget to complete while some others were reluctant to adhere to it at the early stages. Some health workers in India were falling short in their data submissions. The project team has realized that a part of the reason was their hesitation to report the data stemming from a fear that the new technology could be used to monitor their on the job performance. However, with help of the heads of health departments, the misunderstanding was clarified. That, complemented with additional training, rectified the problem.

At the end of the initial two month incubation period, the project team conducted a usability exercise. The aim of the exercise was to assess the health workers' capability to install and configure the mHealthSurvey software, their efficiency in data digitization and submission, and the effectiveness of the standard upkeep procedures of the application. Generally more senior and experienced health workers in India were somewhat slower in completing the exercises, often exhausting the allotted time limits. However, younger research assistants on duty in Sri Lanka were able to complete the exercises well within the benchmark limits.

Finding time to submit data without disrupting current work flows was a significant barrier to adoption of the technology for real-time data submission by the Indian health workers. A large number of them opted to submit the data about 2 – 3 days late, when they had time to spare. This was not the same in Sri Lanka as the only task of the recruited assistants was the data entry. Some of them shadowed medical officers and were able to digitize the data at the same time the patient was examined. Others would collect the treatment chit at the pharmacy when it was submitted for filling the prescribed medication. Yet another group would instead be asked to come in the afternoon when the workload at the hospital was lower, and to process batches of accumulated chits. Occasionally, when the number of patients seen on a given day was particularly large, these assistants would be asked to complete their work on the following day, leading to delays in data submission.

The data entry assistants in Sri Lanka were recruited from primary health centers organized by a highly regarded community organization. Yet, they had a limited knowledge of medical terminology. It was obvious from the results of a controlled experiment which revealed as much as 45% error rates in data entry by some of these assistants. Most of their mistakes were due to inexperience in deciphering often hardly legible handwriting by medical officers. Moreover, the assistants who were not held accountable for fidelity of their submissions had no incentives to maintain good quality. However, more experienced and accountable Indian health workers performed at error rates under 4%.

Impact of Data Quality on Reliability of Event Detection

The utility and performance of T-Cube Web Interface as a statistical surveillance analysis tool largely depends on the quality of the input data. Inconsistencies in the categorical data submitted by the health workers are generally treated as statistical noise by the underlying analytic algorithms. This adversely impacts event detection power and it may lead to increased false alert rates as well as to missing emergence of a real bio-event.

In one instance, TCWI alerted of escalation of whooping cough in Sri Lanka, which was quite unlikely since this disease has been eradicated in the country for several decades. Apparently, in the hurry of entering hundreds of records, the data entry assistant had accidentally mistaken whooping cough for worm infestation. At another occasion in India, possible dengue fever alert caught by TCWI turned out to be a result of a data entry error and it should be identified as an emerging dysentery event instead. The interactive drill-down and visualization functions of TCWI allow the epidemiologists to perform confirmatory analyzes for each alert issued by the automated detection algorithms of TCWI, and therefore to quickly determine their actual validity.

To mitigate the noise, the project defined a set of high priority diseases, symptoms, and signs. All other diseases, symptoms, and signs were categorized commonly as "other". The shortcoming of this approach, although it does reduce the data entry interface complexity and therefore the data entry noise, is that it does not capitalize on the full analytic capabilities of TCWI to find correlations in data when it is clumped in a low-resolution "other" category.

To validate the accuracy of the implemented statistical algorithms, RTBP team extracted publicly available Weekly Epidemiological Returns (WER) data. It tabulates weekly aggregates of notifiable diseases for each of the governing districts in Sri Lanka. The data for years 2008 and 2009 was synthesized to the same shape and form as that being collected through the RTBP mHealthSurvey. The synthesis entails splitting the low granularity district WER data using an assumed probability distribution to reflect the day-of-week visitation densities, as well as patient gender and age group distributions (Shmueli & Burkom, 2010). The data was then processed by TCWI and the results of temporal and spatio-temporal analyzes were compared against the ground truth in WER data.

We found TCWI to produce reliable signals pertaining to several of the news worthy outbreaks of leptospirosis as well as dengue fever events as early as two months ahead of the original time of their detection with use of the pre-existing processes; dengue fever emerging outside of the anticipated rainy season trend cycle; multiple instances of food poisoning cases not detected using the existing processes but emerging in and around the same time and area of a single reported incidence.

The project gathered ground truth data for a span of six months, which was extracted through the various forms and registries maintained by the health departments in the RTBP pilot area. Thereafter, we identified a representative set of significant events from this extraction. TCWI was then put to test to see whether it could detect those events faster than the present system.

Given the nature of the statistical testing based on the magnitude of the frequencies relative to the baseline, TCWI algorithms were able to detect events within approximately 1–2 days after their onset. For example, relative to the baseline data, five cases of viral fever may not seem significant. However, two days later when the counts increase two or three fold, the emerging pattern can be reliably detected. In comparison, the present paper-based communication system requires approximately 4–7 days to just communicate data to the health departments. Therefore, the RTBP approach can be far more efficient.

Over the past several months, TCWI was successfully used in the field to detect actual episodes of escalating Acute Diarrheal Disease, Respiratory Tract Infection, Dysentery, Dengue, and Viral Fever, in the RTBP real-time data, before the health departments came to know about those incidents using the preexisting procedures.

Workability of the Alerting Components

Health departments requested that four SAM templates be created for each combination of the disease category (notifiable, other-communicable) and response categories (action oriented alert and non-action oriented situational awareness message). In addition to the four CAP messaging templates, a fifth template was introduced: weekly list of top 5 diseases. The purpose of this template was to disseminate the top five diseases with the highest counts in that jurisdiction for the given week.

The project team had no difficulties in adopting the Common Alerting Protocol global standard for issuing heath alerts. The medical officers, health inspectors, and nurses receiving the alerts requested that the messages carried the name of the disease, number of involved cases, affected locations, age groups, genders, and the onset. Further to the basic elements of the messages, particularly the action oriented alerts, the recipients were interested in receiving response instructions that were specific to the disease type. The research team was able to map each of the requested pieces of information to a CAP element.

The exercises conducted with the health workers in the efforts to study the completeness of the message had no bearing on the type of technology used; i.e., SMS, Email, or Web. Both Email and Web alerts were capable of carrying a complete message with all attributes but the SMS was constrained by its content size limitations. However, evidence from the exercise shows that all health workers were able to decipher the messages with absolute certainty.

The exercises exposed some dilemmas due to technical faults, sometimes leading to repetitions of the procedure of message creation and delivery. Most such repetitions were caused by weak Internet connectivity. The users participating in the usability exercise showed a high level of competency.

The health departments in Sri Lanka were quick to appreciate the efficiencies and benefits of electronic messaging when compared to the current paper-and-mail-based system. They are now able to communicate all significant events detected through TCWI to health workers in the affected areas using the CAP compliant SAM software.

CONCLUSION

The present Indian and Sri Lankan paper-based disease notification programs invest bulk of their resources in data collection activities, sparing very little or no emphasis on event detection and alerting. These systems are passive in nature, they do not offer prognostic capabilities, and serve primarily as reporting facilities for counting disease cases. Collecting data in the current way is costly and time consuming, it requires substantial resources for manual processing of data with multiple stages of aggregation at various administrative levels. Operating costs of running the Indian system's collection, detection, and alerting capabilities exceeds the projected cost of running RTBP in the analogous environment by the factor of two. The cost comparison in the case of Sri Lankan system is not as spectacular, but it is still significantly in favor of RTBP. In both cases, human resources currently allocated to manual processing of data, after implementation of RTBP could be assigned to other duties such as participation in crisis responses or preventative activities, making public health departments more resourceful when and where it matters most.

The RTBP is not just less expensive to operate. It is also far more efficient, providing timely and comprehensive coverage of the public health status at a much greater detail than the existing systems. It enables a proactive approach to public health management with the introduction of rapid detection and efficient alerting functionality. Better situational awareness, attainable with the use of RTBP, will allow public health managers to stage responses to emerging crises much earlier and in a more focused manner than currently possible. That would provide substantial tangible and intangible, societal and budgetary benefits to individuals and to the local and national administrations. More effective responses can save lives, protect vulnerable populations, and save substantial amounts of money and other resources.

The data collected through RTBP is substantially more detailed and comprehensive than the data collected using the current systems. Syndromic information contained in RTBP data will allow for detecting emergence of diseases which manifest with unique sets of symptoms and signs, even before the correct diagnoses are known to medical community. RTBP data also reaches beyond the domain of a small number of notifiable or even communicable diseases. It records cases of diabetes or arthritis as well setting up the stage for future research into prevalence of such societally important ailments, and letting public health managers to monitor trends, to make effective plans of their treatment as well as plans for allocation of required resources.

RTBP is also affordable to implement and feasible to maintain in developing countries. Its field data collection infrastructure relies on an inexpensive cell phone technology which also has a ubiquitous coverage and technical support even in rural areas of India and Sri Lanka. The analytic software can be accessed using common purpose personal computers with Internet connectivity and a web browser, which already exist at regional health departments. A health department's expenditure for the event detection and response component of the RTBP is the same compared to the present system because they already have such infrastructure in place. Any licensing costs of the software can be distributed across multiple districts and health departments, making a single location cost share nearly negligible.

RTBP offers a cost effective and operationally efficient alternative to the existing predominantly paper-based disease reporting systems in many developing countries. It takes advantage of new Information and Communication Technology to improve awareness of public health situation in those vulnerable regions. The results of the ongoing pilot studies conducted in India and Sri Lanka indicate that the health workers in these countries are fully capable of adopting the new technology for disease surveillance, and that the policy makers are willing to consider implementation of such solution to overhaul the existing processes to benefit their populations. We believe that the RTPB technology is highly applicable in a range of environments, beyond monitoring public health. We are also looking forward to scaling it up in India and Sri Lanka as well as to testing it in other developing countries.

ACKNOWLEDGMENT

This work was carried out with the aid of grant 105130 from the International Development Research Center, Ottawa, Canada, and partially supported by the US National Science Foundation, award 0911032.

REFERENCES

ADD/Cholera statistics (2010). Epidemics, *Department of Public Health and Preventive Medicine, Health and Family Welfare Department, Tamil Nadu, India*. Retrieved from http://www.tnhealth.org/dphdb.htm

Agampodi, S., Somaratne, P., Priyantha, M., & Peter, M. (2008). *An interim report of Leptospriosis outbreak in Sri Lanka*. Epidemiology Unit of Sri Lanka.

Ammenwerth, E., Jytte Brender, J., Pirkko Nykänen, P., Prokosch, H.-U., Rigby, M., & Talmon, T. (2004). Visions and strategies to improve evaluation of health information systems Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. *International Journal of Medical Informatics, Elsevier publications, 73*, 479-491.

Anderson, J., & Aydin, C. (2005). *Evaluating the organizational impact in healthcare information systems, Second edition, Health Informatics Series*. Berlin: Springer.

Careem, M., De Silva, C., De Silva, R., Raschid, L., & Weerawarana, S. (2006). Sahana: Overview of a Disaster Management System. In *Proceedings of the International Conference on Information and Automation*, Colombo, Sri Lanka.

Chapman, R. (2009). *Public Health Information Network (PHIN) Communication and Alerting Guide (PCA), version 1.1*. US Center for Disease Control.

DHF Report – Cases & Deaths. (2010). *Distribution of Notified (H399) DF/DHF Cases 2010, Epidemiology Unit, Sri Lanka*. Retrieved from http://www.epid.gov.lk/Dengue_updates.htm

Dubrawski, A. (2010). Detection of Events in Multiple Streams of Surveillance Data. In Zeng, D., Chen, H., Castillo-Chavez, C., Lober, W., & Thurmond, M. (Eds.), *Infectious Disease Informatics and Biosurveillance*. Berlin: Springer Verlag.

Friedman, C., & Wyatt, J. (2006). *Evaluation methods in Bioinformatics, second edition, Health Informatics Series*. Berlin: Springer.

Ganesan, M., Prashant, S., Janakiraman, N., & Waidayanatha, N. (2010). *Real-time Biosurveillance Program: Field Experiences from Tamil Nadu, India. IASSH conference paper, Varanasi*. India: Uttarpradesh.

Gow, G., Vincy, P., & Waidyanatha, N. (2010). Using mobile phones in a real-time biosurveillance program: Lessons from the frontlines in Sri Lanka and India. In *Proceedings of the IEEE International Symposium on Technology and Society (ISTAS '10)*, Wollongong, New South Wales, Australia.

Kannan, T., Sheebha, R., Vincy, A., & Waidyanatha, N. (2010). Robustness of the mHealthSurvey midlet for Real-Time Biosurveillance. In *Proceedings of the 4 IEEE International Symposium on Medical Informatics and Communication Technology (ISMICT '10)*, Taipei, Taiwan.

Lewis, D. (2003). Evaluation of Public Health Informatics. Public Health Informatics and Information Systems. In P. O'Carroll, W. Yasnoff, M. Ward, M., & L. Ripp (Eds.), *Health Informatics Series* (pp. 239-266). New York: Springer.

Neill, D., & Moore, A. (2006). Methods for Detecting Spatial and Spatio-Temporal Clusters. In Wagner, M., Moore, A., & Aryel, R. (Eds.), *Handbook of Biosurveillance* (pp. 243–254). London: Elsevier Academic Press. doi:10.1016/B978-012369378-5/50018-1

Neill, D., & Cooper, G. (2010). A Multivariate Bayesian Scan Statistic for Early Event Detection and Characterization. *Machine Learning*, *79*, 261–282. doi:10.1007/s10994-009-5144-4

Prashant, S., & Waidyanatha, N. (2010). User requirements towards a biosurveillance program. In Kass-Hout, T., & Zhang, X. (Eds.), *Biosurveillance: Methods and Case Studies* (pp. 240–263). Boca Raton, FL: Taylor & Francis.

Ray, S., Michalska, A., Sabhnani, M., Dubrawski, A., Baysek, M., Chen, L., & Ostlund, J. (2008). T-Cube Web Interface: A Tool for Immediate Visualization, Interactive Manipulation and Analysis of Large Sets of Multivariate Time Series. In *Proceedings of the AMIA Annual Symposium*, Washington, DC (p. 1106).

Sampath, W., Waidyanatha, N., Ariyaratne, V., Ratnayake, R., Hemachandra, P., & Edirisinghe, E. (2010). Real-Time Biosurveillance Pilot Program. *Sri Lanka Journal of Bio-Medical Informatics*, *1*(3), 139–154. doi:10.4038/sljbmi.v1i3.1774

Shmueli, G., & Burkom, H. (2010). Statistical Challenges facing early outbreak detection in Biosurveillance. *Technometrics*, *52*(1), 39–51. doi:10.1198/TECH.2010.06134

Siriwardana, C. (2008, August 12). Sri Lanka Kidney disease epidemic leaves doctors baffled. *World wide web news article, Science and Development Network*.

Sub-Committee of Technical Experts on Clinical Management of DF/DHF. (2005). *Guidelines on Clinical Management of Dengue / Dengue Haemorrhagic Fever (2005), Epidemiological Unit, Ministry of Health, Sri Lanka*. Retrieved from http://tinyurl.com/289vkr9

Vital Wave Consulting. (2009). *mHealth for Development: The Opportunity of Mobile Technology for Healthcare in the Developing World*. Paper presented at the UN Foundation-Vodafone Foundation Partnership.

Wagner, M. (2006). Challenges of Biosurveillance: Introduction. In Wagner, M., Moore, A., & Aryel, R. (Eds.), *Handbook of Biosurveillance* (pp. 3–12). London: Elsevier Academic Press. doi:10.1016/B978-012369378-5/50003-X

Nuwan Waidyanatha is a Senior Researcher with LIRNEasia and is the Project Director of the research - "Evaluating a Real-Time Biosurveillance Program". His expertise are in ICT system design and actionable research working collaboratively with multiple national and international stakeholders. He applies the academic and industrial experience in providing practical solutions for developing nations. Present research interests are in developing early warning systems and information exchange platforms for emergency communication. He is an advocate of Free and Open Source Software and is a member of the Project Management Committee of the Sahana Software Foundation. His master's degree is in Operations Research and bachelor's degree in Mathematics and Computer Engineering.

Artur Dubrawski researches intelligent systems that work, are useful, and make economic sense, and ways to effectively build and deploy them. For over 20 years now his work has been driven by real-world applications, currently in the areas of public health, food safety, nuclear safety, health of equipment, and predictive maintenance. It involves investigating machine learning algorithms and data structures to facilitate probabilistic modeling, predictive analysis, interactive exploration, and understanding of data. Dr. Dubrawski directs the Auton Lab and teaches graduate courses on data mining and business intelligence at Carnegie Mellon University. He leads fundamental and applied research projects funded by government agencies and industrial partners. Artur Dubrawski received a doctorate in robotics and automation from Polish Academy of Sciences, and a M.Sc. in aircraft engineering from Warsaw University of Technology. He spent a substantial part of his professional career serving in managerial and technical roles in high-tech industry.

Ganesan M has a Ph.D in Energy from the School of Bio-Energy and his Masters in Rural Development from Madurai Kamaraj University. He has worked in the sustainable rural development projects with various reputed Non Governmental and Governmental organizations for 15 years. He joined as Core Trainer under Vazhndhu Kaatuvom project funded by World Bank. Since July 2007, he has been involved in the health projects with IITM's Rural Technology and Business Incubator. His interest includes exploring the use of ICT tools in facilitating service delivery to rural areas.

Gordon Gow is an Associate Professor in Communication and Technology at the University of Alberta. His teaching and research examines the impact of new communication technology on public health and safety. His areas of expertise include mobile and wireless communications, telecommunications policy, Internet and social media. Currently, he is principle investigator on a multi-year SSHRC-funded project (Standard Research Grant) to examine campus alerting in Canada. Dr. Gow is also director of the MARS micro-lab (Mobile Applications for Research Support) at the University of Alberta. The MARS lab provides support for projects investigating the use of the mobile communication devices in support of scholarly and community-based social science research.