A PUBLIC HEALTH APPLICATION OF DATA ANALYSIS FOR HOMELAND SECURITY

Marjorie Greene
Robert Eek

Science Applications International Corporation
1710 SAIC Drive
McLean, VA 22102, U.S.A.

ABSTRACT

This presentation follows up the talk last year to WINFORMS (The Washington Institute for Operations Research and The Management Sciences) in which an approach developed for the analysis of military command and control during crises was shown to be relevant to surveillance for an infectious disease outbreak. ProMEDMail is an Internet-based system dedicated to rapid global dissemination of information on outbreaks of infectious diseases and acute exposures to toxins that affect human health. The presentation will demonstrate how the collection, formatting, and analysis of raw data, using ProMED-Mail, can point to emerging biological incidents and allow the real-time dissemination of results to local, regional, and central health facilities.

1 INTRODUCTION

The medical community needs the capability to model, simulate, and wargame biological threats such as pandemics and bioterrorism that may impact homeland security. Epidemiological information needed to identify pandemics or bioterrorism exists in many sources, but is not presented in a form suitable for alerting or doing records-based analysis. Needed information exists predominantly in unstructured form and requires extensive manipulation to be usable. Due to this lack of structure, analytical activities related to biological threats are ad hoc, non-real time and non-reproducible.

There is also increasing recognition that the manifestations of bioterrorism and biowarfare may be difficult to distinguish from naturally arising infections. Natural emerging infections are appearing more frequently and assuming forms that presage ever-increasing dangers for the health of populations worldwide. This, in turn, poses dangers for homeland security. Major programs have been initiated to understand the nature of viruses and to develop laboratory tests for their identification. *More and more, however, there is recognition that the first, and most critical, line of defense must be an integrated worldwide detection and warning system for emerging infectious diseases. Such a system will ultimately have to pull information from a variety of sources, including open-source population health information, and make those data available in interestingly useful ways using appropriate analysis and visualization tools.*

A number of organizations and countries have established computer networks to facilitate the reporting of disease detections. They are often characterized by significant time delays in propagating critical reports. Some are strictly focused on a subset of diseases. And none is coupled with capabilities for effectively collating and formatting the data, analyzing the information, and quickly disseminating the results to local, regional, and central health facilities.

2 APPROACH

Figure 1 illustrates the implementation of a monitoring system to provide decision support and medical management tools to the U.S. for homeland security. Through an enhancement of ProMED-Mail, the capability to use unstructured text to point to a biological event is demonstrated. ProMED-Mail is an Internet-based system dedicated to rapid global dissemination of information on infectious diseases and acute exposures to toxins. It has over 30,000 subscribers (military, civilian) in more than 150 countries. ProMED-Mail is a major source of expertise and information. It is an official program of the International Society of Infectious Diseases. Its users include WHO, UN, Red Cross, CDC, and others. It has the largest reporting base of any health organization.
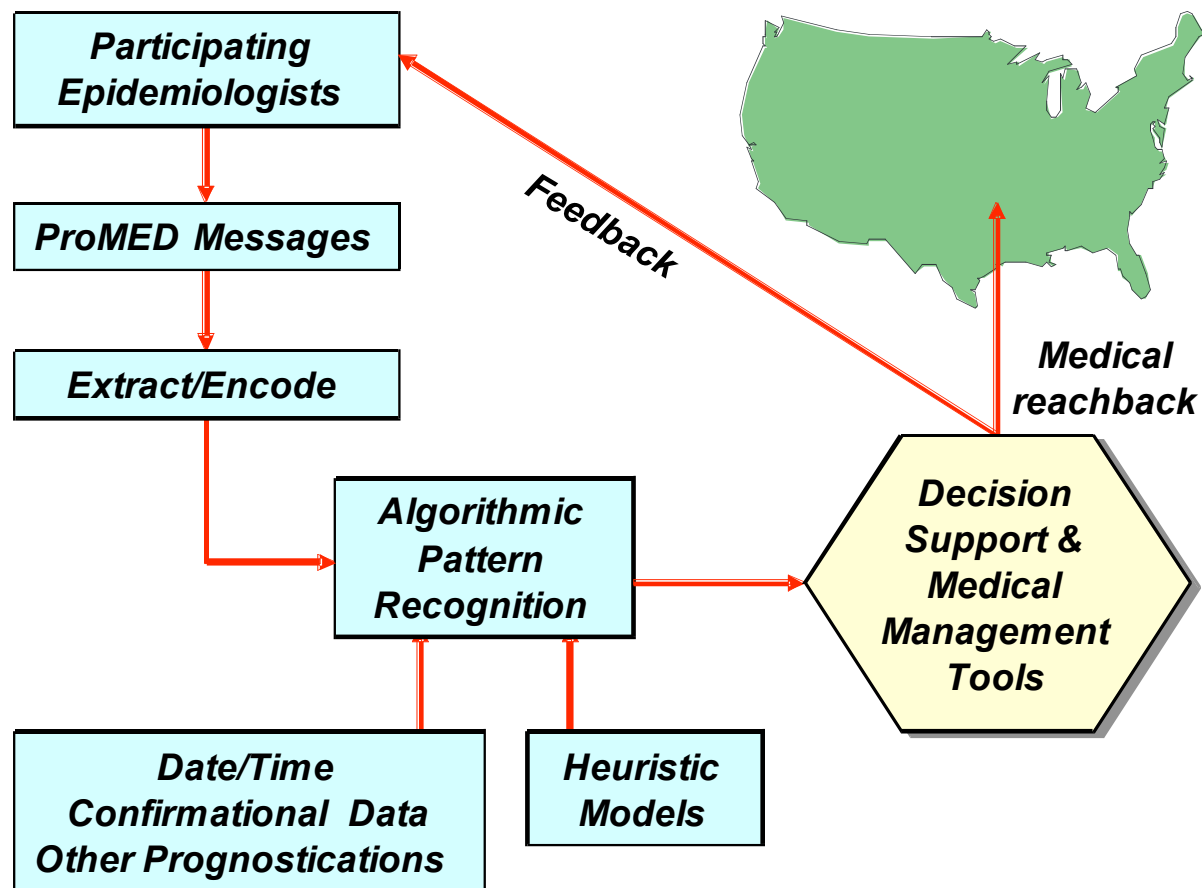
Figure 1: Enhancing ProMED-mail for Homeland Security

The system we are building is not the only surveillance system available to homeland security. Rather, it is a complementary source of information that provides an early detection capability with minimal false alarms, under conditions of severe noise and high degrees of uncertainty. It builds on an infrastructure already established to access specific subsets of enterprise-wide data in ProMED-Mail and tailors it to the early detection of infectious diseases. It does this by designing a database and adding another capability – the development of algorithms to automatically identify patterns that point toward a potential outbreak. Ultimately, our system provides analyses, presentation, and visualization services to epidemiologists and other appropriate users by building on ProMED-Mail and providing the following enhancements:

- Automated extraction and formatting of data.
- Expert systems for epidemiologic analysis.
- Powerful algorithms for data and information fusion.
- Interconnection with other military and civilian medical surveillance systems.

3 METHODOLOGY

After input data are extracted, formatted and organized, they are put into a series of algorithms for the purpose of information fusion. Reports are correlated with other recent events/reports from this and other systems. The system complements the already well-established system of the DoD Global Emerging Infection Surveillance System (DoD-GEIS) and the alerting algorithms provided by the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE).

Patterns are matched against existing templates. Using statistical and heuristic approaches, we determine the validity and significance of the data. The system identifies agents (**including never-before-seen agents**) and checks for redundancies and inconsistencies. It will characterize the symptoms, the causative agent, and the contribution environment. Data are tabulated and analyzed with algorithms that guard against false alarms yet provide early detections. Eventually, a software interface tool will also be produced to search and access the data for other purposes, such as research-oriented work. The database data are con-

tinually updated and adapted to new input information, automatically, so as to improve our knowledge of agents, diseases, hosts, and the environment, over time.

4   RESULTS

We define an operational concept for mining ProMED-Mail and other unstructured sources for information useful in the detection of events related to emerging infectious diseases. Our immediate goal was to determine what characteristics an automated medical surveillance system with utility to homeland security should possess. Our next step will be to design and implement a pilot analytical system that can be demonstrated in supporting first responders, local medical facilities, and other players to respond to biological threats such as pandemics and bioterrorism.

AUTHOR BIOGRAPHIES

MARJORIE GREENE is a senior program manager with SAIC currently supporting the Office of Naval Research. She has over 30 years of experience in strategic planning, analysis and Naval Operations. She has experience in DoD, Navy, Air Force, commercial, state and federal government programs, and other database and applications research, design, and development, C4I, and information systems requirements analysis. She has subject matter expertise in the telecommunications, financial, and health care industries. As a member of the Management Team of the HUBS (Hospitals, Universities, Businesses, Schools) program, Ms. Greene developed offerings in middleware, XML, and other reusable software solution sets for enterprise customers. She initiated a project to enhance state health information networks as the foundation for rapid detection and reporting of outbreaks of illness resulting from bioterrorism. Ms Greene conducted needs assessments for public health initiatives designed to improve the effectiveness of information gathering. She developed approaches to track outcomes and consequences of the use of communications technology for health services. Earlier, Ms. Greene developed analytical tools for describing information flow throughout the Naval chain-of-command.

ROBERT EEK is a principal engineer and program manager at SAIC with over 30 years experience in the design and implementation of complex information systems for the U.S. Navy, DoD Agencies, State Government, and other organizations. He has experience with database and applications research, design and development, C4I and information systems requirements analysis, software development and secure systems engineering. He designed the data archive for the University of Hawaii Institute for Astronomy's Panoramic Survey Telescope and Rapid Response System (Pan-STARRS). Earlier he designed and implemented a business intelligence database for NGA that combined source holdings and job cost/schedule information with geospatial search support. Dr. Eek helped reengineer the 300-gigabyte Population Health metrics (PHOTO) data mart for the Military Health Systems. For DISA, he supported the DII Common Operating Environment (COE) Architecture Oversight Group, Management Services Working Group and Data Access Services Working Group. Dr. Eek co-authored the *DII COE Integration and Run-Time Specification (I&RTS)* and wrote the database and DBMS integration standards and segmentation policies of that document.