SAGES Overview: Open-Source Software Tools for Electronic Disease Surveillance in Resource-Limited Settings

Brian H. Feighner, Timothy C. Campbell, Aaron T. Katz, Richard A. Wojcik, Jacqueline S. Coberly, Shraddha V. Patel, Wayne A. Loschen, and Sheri H. Lewis

Disease surveillance, the foundation of public health practice, is undergoing a revolution driven by advances in information technology. The past 15 years have seen vast improvements in the collection, analysis, visualization, and reporting of public health data. Resource-limited countries have lagged behind because of challenges in information technology infrastructure and public health resources. The Suite for Automated Global Electronic bioSurveillance (SAGES) is a collection of modular, open-source software tools designed to meet the challenges of electronic disease surveillance in resource-limited settings. Individual SAGES tools may be used in concert with existing surveillance applications or en masse for an end-to-end biosurveillance capability. This flexibility allows for the development of an inexpensive, customized, and sustainable disease surveillance system. The ability to rapidly assess anomalous disease activity may lead to more efficient use of limited resources and better compliance with World Health Organization International Health Regulations.

INTRODUCTION

Disease surveillance was defined by Langmuir in 1963 as "the continued watchfulness over the distribution and trends of incidence through the systematic collection, consolidation and evaluation of morbidity and mortality reports and other relevant data," with the "regular dissemination of the basic data and interpretations to all who have contributed and to all others who need to know."[1] Thacker expanded and refined this definition in 1988 when he wrote, "Public health surveillance is the ongoing systematic collection, analysis and interpretation of outcome-specific data for use in the planning, implementation, and evaluation of public health practice."[2] This linking of public health surveillance with the evaluation of public health practice emphasizes its primary purpose—to direct the expenditure of limited public health resources in a manner that yields the greatest return on investment.[3]

Public health surveillance involves clearly defining events of public health interest, counting those events, and then analyzing those events with respect to person, place, and time. For example, the U.S. Centers for Disease Control and Prevention case definition for an influenza-like illness is "fever (temperature of 100°F [37.8°C] or greater) and a cough and/or sore throat without a KNOWN cause other than influenza."[4] Patients meeting the case definition of influenza-like illness are counted by sex and age category (person) and characterized by site (place) and date of onset (time). This conceptually simple process not only characterizes the level, distribution, and spread of influenza-like illness in the community but also suggests useful information such as determinants of disease transmission, possible mitigating strategies, and future prevention strategies. Public health surveillance may be performed on all patients, so-called universal surveillance, or performed at designated sites felt to be representative of the population as a whole, so-called sentinel surveillance. Surveillance may also be described as active, when public health officials contact health care providers, or passive, when public health officials rely on reports from health care providers. A wide variety of data sources are used in public health surveillance, including vital statistics, health reports, hospital records, laboratory reports, outpatient visits, registries, and health surveys.

Disease surveillance is commonly recognized for its ability to detect disease outbreaks. Simply put, unless the baseline level of disease is well understood, it is difficult to identify disease levels significantly in excess of normal.[2] This is an important function, and the early detection of anomalous disease events, particularly the intentional release of pathogens, has received much recent attention. Critics point out that disease surveillance, particularly syndromic surveillance, may not catch small outbreaks of disease that remain hidden in the background noise and also note that diseases with shocking presentations, such as hemorrhagic fevers, are generally identified by health care providers.[5] Nevertheless, disease surveillance plays a critical role in the detection of disease outbreaks.[1–3,6,7] Importantly, in the case of small- to medium-size outbreaks distributed over a wide geographic area—now common because of large, centralized food-processing plants—coordinated disease surveillance identifies problems that might otherwise go unnoticed in each local jurisdiction.[6] Disease surveillance accomplishes several additional important functions to direct the sage practice of public health.[2] Disease surveillance identifies and quantifies the diseases most burdensome to a population. How a disease spreads through the population of interest and how it affects individuals over time can both be carefully documented by disease surveillance. Importantly, disease surveillance is used to evaluate public health interventions, identifying effective and ineffective public health practices. Disease surveillance can suggest hypotheses, direct research, and detect changes in the practice of clinical (or veterinary) medicine over time. Effective disease surveillance, though not always exciting, is the foundation of successful public health practice.

For centuries, disease surveillance was a paper-based process. In the 1990s, with the emergence of inexpensive, powerful information technology tools, disease surveillance became an electronic process in wealthy countries.[7] Incorporating information technology advances led to startling improvements in the timeliness of reporting and the sophistication of data analysis. Such systems have become versatile tools in health departments in the United States, and electronic disease surveillance holds promise to improve health security in resource-limited environments.[8–12] Epidemiologists using electronic disease surveillance not only have the potential to detect anomalous disease activity earlier than those using traditional laboratory-based surveillance, but they also have the ability to monitor the health of their community in the face of a known threat.[10–12] More than a decade ago, in collaboration with the DoD, the Johns Hopkins University Applied Physics Laboratory (APL) developed the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE).[12] ESSENCE is currently being used by the DoD, Veterans Health Administration, and numerous state and local health departments in the United States.[12] ESSENCE allows for essentially real-time data collection, leading to timely anomaly detection with dynamic, sophisticated data analysis, visualization, and reporting. In addition, electronic disease surveillance systems are able to automatically ingest large amounts of preexisting electronic data streams for analysis. These data sources, such as insurance claims, pharmaceutical data, and commercial sales, differ from traditional sources, such as health data from medical treatment facilities, yet they often have content that is relevant to public health.[12] The current Suite for Automated Global Electronic bioSurveillance (SAGES) initiative leverages the experience gained in the development of ESSENCE, and the analysis and visualization components of SAGES are built with the same features in mind.

Emerging and reemerging infectious diseases are among the most serious threats to global public health.[13,14] The World Health Organization (WHO) has identified more than 1100 epidemic events worldwide in the last 10 years alone.[15] The emergence of the novel 2009 influenza A (H1N1) virus and the SARS coronavirus in 2002 has demonstrated how rapidly pathogens can spread worldwide.[13–16] This infectious disease threat, combined with a concern over man-made biological or chemical events, spurred WHO to update its International Health Regulations (IHR) in 2005.[17] The new 2005 IHR, a legally binding instrument for all 194 WHO-member countries, significantly expanded the scope of reportable conditions and are intended to help prevent and respond to global public health threats. Specifically, the IHR require strengthening disease detection and response capacities in order to report, within 24 hours of assessment, any public health event of international concern. SAGES, an electronic biosurveillance initiative described herein, aims to improve local public health surveillance and IHR compliance, with particular emphasis on resource-limited settings.

SAGES AND ITS TOOLS

The U.S. Armed Forces Health Surveillance Center is committed to enhancing electronic disease surveillance capacity in resource-limited settings around the world. In 2008, its Global Emerging Infections Surveillance and Response System Division (AFHSC-GEIS) entered into a robust collaboration with the APL to create SAGES. Aware of the work of others underway on individual surveillance systems components (e.g., collection of data by cell phones), we focused our efforts on the integration of inexpensive, interoperable disease surveillance software tools that facilitate regional public health collaborations.[18]

SAGES tools are organized into four categories: (*i*) data collection, (*ii*) analysis and visualization, (*iii*) communications, and (*iv*) modeling/simulation/evaluation (Fig. 1). Within each category, SAGES offers a variety of tools compatible with surveillance needs and different types or levels of information-technology infrastructure. In addition to offering flexibility in their selection, the tools also do not require a fixed database format. For example, rather than requiring an existing database to adapt to the tool, the SAGES database tools adapt to all Java database–compliant formats, a mainstay of enterprise information-technology development for decades. Lastly, the SAGES tools are built in a modular nature, which allows the user to select one or more tools to enhance an existing surveillance system or to use the tools en masse for an end-to-end electronic disease surveillance capability. Thus, each locality can select tools from SAGES on the basis of their needs, capabilities, and existing systems to create a customized electronic disease surveillance system.

Data Acquisition

Rapid data acquisition is arguably the most challenging aspect of establishing a successful electronic disease surveillance system.[10–12] In resource-limited settings, it is not unusual for disease surveillance to be paper based, with data sent to upper echelons only when a courier happens to be going that way. This system of data acquisition often results in a data lag—the time from health event to analysis—of weeks to months. Because the ability to mitigate the effects of an infectious disease outbreak strongly depends on early recognition and response, minimizing data lag is of utmost importance. In resource-limited settings, it is imperative to select technologies that are both easy to incorporate into existing health services and sustainable with little or no additional financial investment. The approach should allow customizable data collection, enable multiple data streams collected in different ways, and be scalable on the basis of needs.[8,9] Novel data collection tools included within SAGES are web forms, short message service (SMS) texting using simple cell phones, forms on Android smartphones transmitted by SMS, digital logbooks, and forms on tablets using Wi-Fi systems. Where appropriate, other collection methods such as e-mail and secure file transfer protocol can be applied as well.

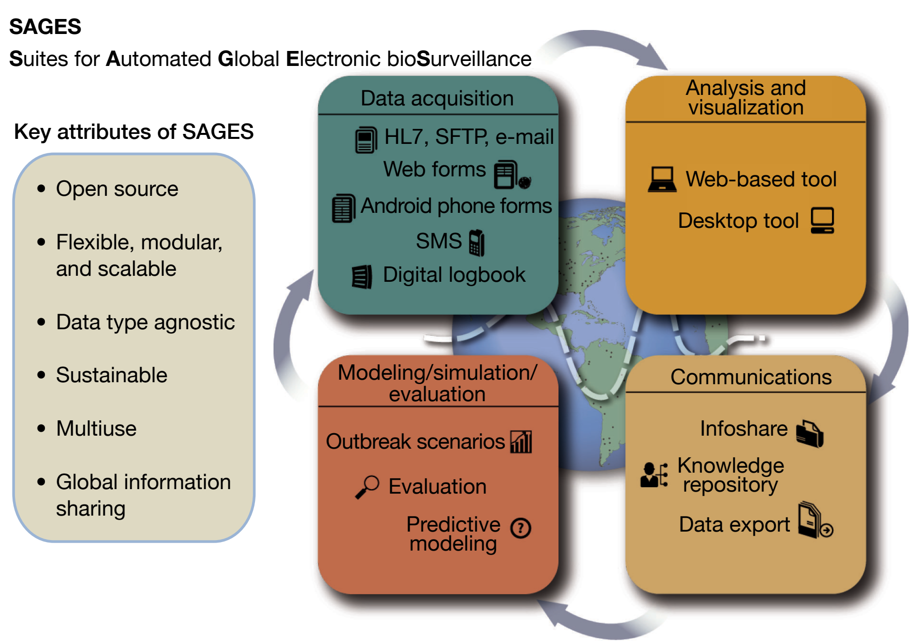

Figure 1. The four types of SAGES software tools.

Analysis and Visualization

As previously discussed, the SAGES analysis and visualization tools are built on the features and functionality of the mature ESSENCE system. The enterprise ESSENCE system requires a high-speed Internet connection, relies on automated data streams, and uses proprietary software for the display of data.[12] The SAGES web-based application and the desktop application are both open-source tools that provide similar functionality to countries with limited public health resources. All ESSENCE applications contain alerting algorithms developed by both APL and the U.S. Centers for Disease Control and Prevention to identify anomalous events. Users also have the ability to add additional algorithms as desired.

The first SAGES analysis and visualization tool, known as ESSENCE Desktop Edition (EDE), is a stand-alone analysis and visualization tool that can be installed on most computers. EDE does not need access to the Internet because it ingests data files stored on the computer on which it is running. Although it can be run on the most basic computers, EDE allows for extensive epidemiologic evaluation of health data. The flagship tool of the SAGES program is OpenESSENCE, the web-based analysis and visualization tool. OpenESSENCE can be run as a stand-alone application but is most powerful when used with network access. OpenESSENCE enables epidemiologists to monitor the population's health from any computer connected to that network. Available analyses for both EDE and OpenESSENCE depend on the nature of the data ingestion but generally include demographic characterizations, temporal and spatial analyses, display of patient-level information, geographic information system mapping, anomalous event detection, and dynamic query capability (Figs. 2–4). OpenESSENCE also allows surveillance system administrators to monitor data reporting and data lag.

Communications

SAGES tools can facilitate compliance with 2005 IHR reporting requirements and allow the sharing of actionable information across jurisdictional boundaries. Sharing of patient-level data across regional boundaries is generally not realistic and often not helpful because local public health entities are usually best suited to interpret local events. Once the data have been transformed into actionable information, however, it may be immensely valuable to share that information with other countries in the region. Dissemination of this type of information may aid in the interpretation of regional events and helps foster better, lasting public health collaborations. SAGES data visualization and reporting products are exportable into common image formats. Planned data-sharing tools will allow each SAGES user to control the type and level of detail of information shared with each recipient ("role-based access") and also whether the information sharing is manual or automated.[19] This capability does not compete with the WHO Global Outbreak Alert & Response Network (GOARN) but compliments it for organizations, such as Ministries of Defense, that wish to communicate

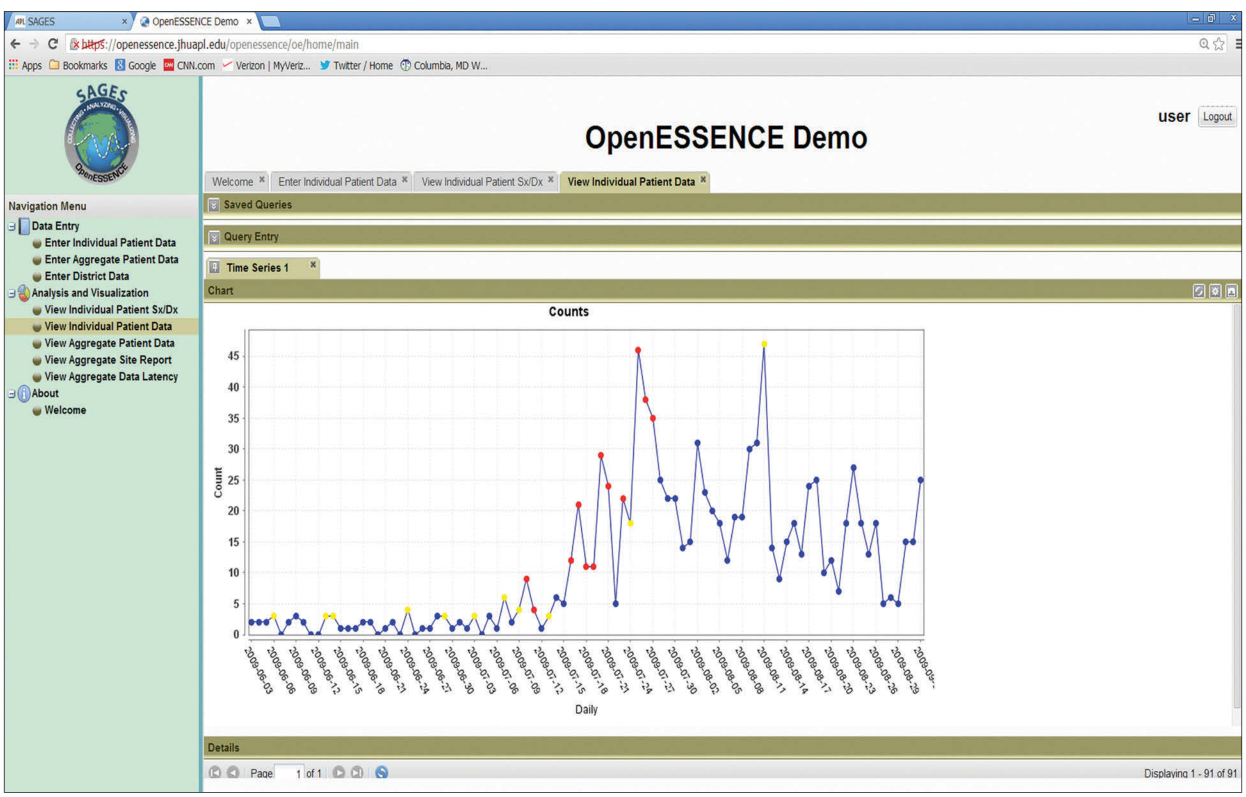

Figure 2. Time series (number of events by day) display in OpenESSENCE, with alerting to anomaly detection (synthetic data).

within or outside of GOARN. APL SAGES personnel and AFHSC-GEIS are currently in discussions with the WHO and WHO regional offices over the use of SAGES tools in support of WHO initiatives. Lastly, and importantly, the data collected using SAGES software remain under the sole control of the user at all times.

Modeling/Simulation/Evaluation

The APL SAGES program has been active in modeling infectious diseases for the U.S. military. The Pandemic Influenza Policy Model (PIPM), an early proprietary software project, allowed military public health officers and planners to vary attributes of respiratory

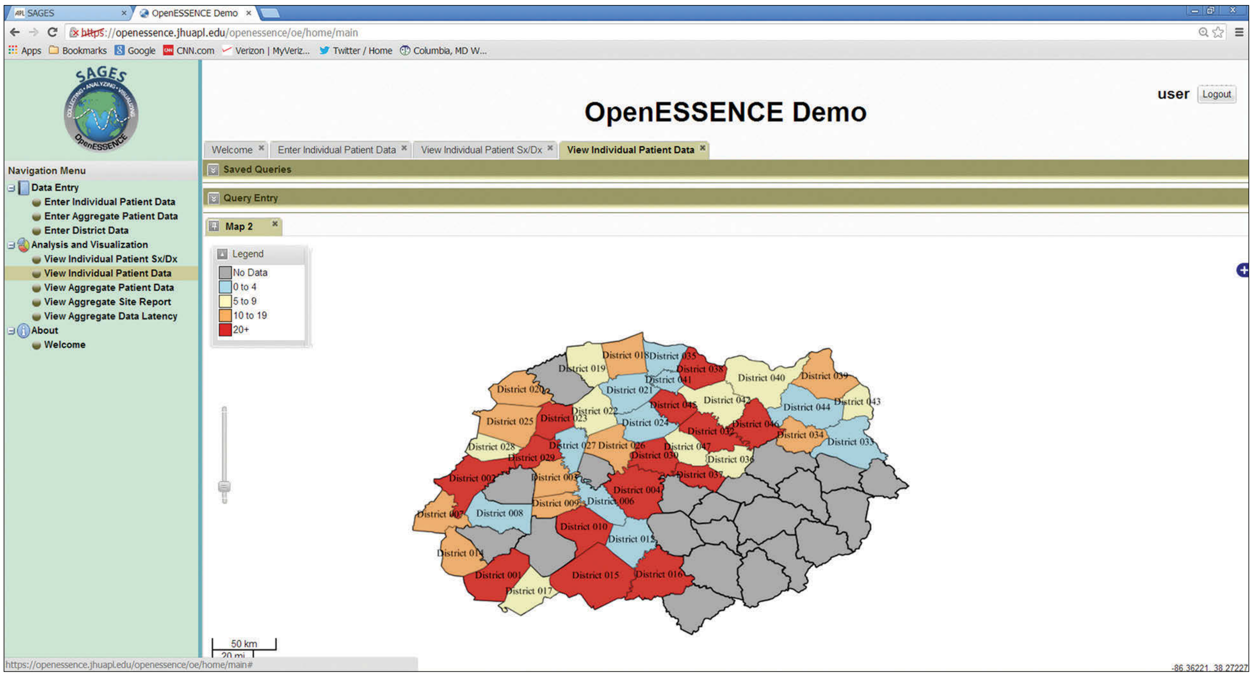

Figure 3. Chloropeth map with level of disease activity indicated by color and district (synthetic data).

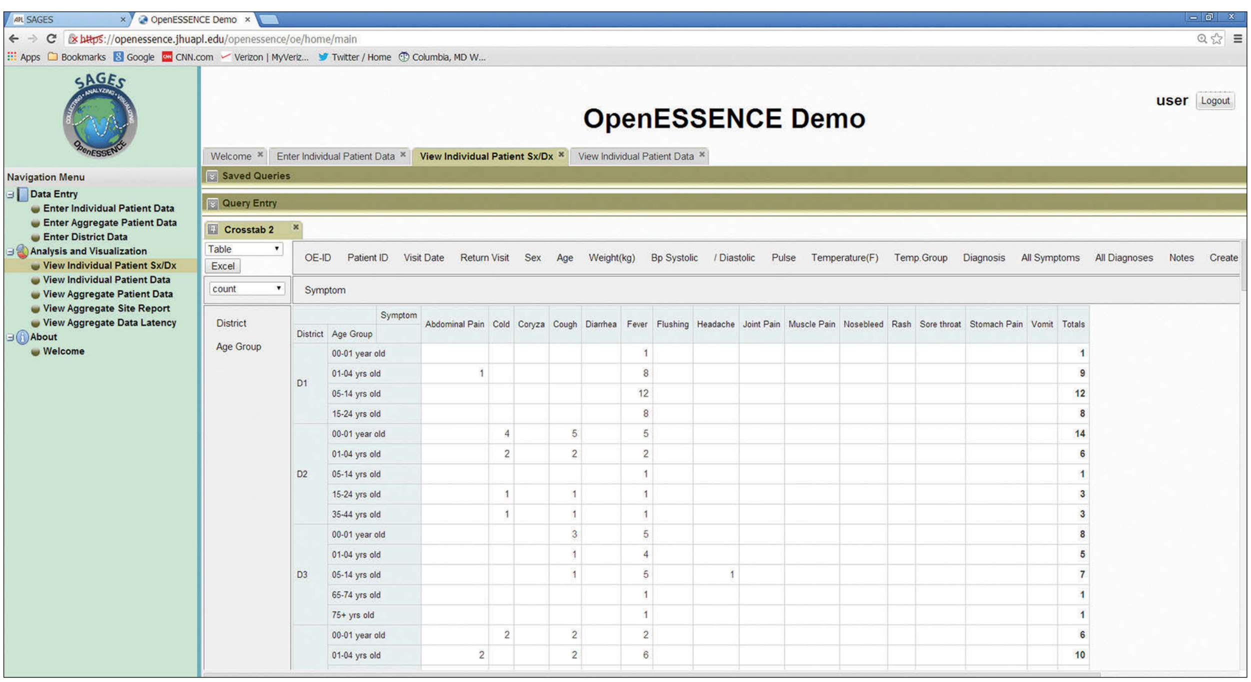

Figure 4. Dynamic generation of data tables for analysis and reporting (synthetic data).

pathogens such as virulence, transmissibility, and susceptibility while also varying public health interventions.[20] Once configured by the user, the PIPM then modeled the transmission of disease from individual to individual in what is termed an "agent-based" model. Using serial PIPM model runs, users could evaluate the effect of different interventions, as well as the effect of the timing of those interventions (e.g., the effect of closing schools upon the spread of influenza on a military base).

APL has sponsored several exercises to train users, test surveillance system features, and evaluate outbreak response. We have developed a number of methods for developing simulated outbreaks, both natural and man-made, which can then be "injected" into a simulated surveillance database for exercise purposes.[21] Past exercises have included simulations of naturally occurring diseases such as pandemic influenza and deliberate release of biological warfare agents such as anthrax, plague, and tularemia. Participants in these exercises have ranged from local health departments to combined (international) military forces.

In addition to agent-based infectious disease modeling, the SAGES program has other techniques for predictive disease modeling. The PRedicting Infectious Disease Scalable Model (PRISM) is the most recently developed proprietary disease prediction tool.[22] PRISM is not an agent-based model but rather uses novel data-mining techniques to predict the future incidence of disease. Early results have demonstrated impressive validity with dengue fever in the Republic of Peru and malaria in the Republic of Korea. It is our desire to explore possible collaborations among our disease modelers and users of electronic disease surveillance systems such as SAGES, with an ultimate goal of using disease surveillance data to improve future disease prediction efforts.

Only those diseases with the highest burdens on the population should be followed, and surveillance systems should be periodically evaluated to determine their usefulness.[1,2,7] At this time, APL is working with AFHSC-GEIS to develop tools for the evaluation of SAGES surveillance systems. Several authors have described the evaluation of electronic disease surveillance systems.[23–27] We intend to perform such evaluations of SAGES surveillance systems and investigate the feasibility of developing automated evaluation tools in the future.

CONCLUSION

Effective and efficient disease surveillance is critical for the economic development of nations. Globalization, particularly rapid international travel, has led to the maxim, "A public health emergency anywhere is a public health emergency everywhere." The SAGES project is intended to enhance electronic disease surveillance capacity in resource-limited settings around the world, hopefully leading to economic development in resource-limited nations and improved global health. We have combined electronic disease surveillance tools developed at APL with other open-source, interoperable software tools to create SAGES. We believe this suite of tools will facilitate local and regional electronic disease surveillance, regional public health collaborations, and international disease reporting. SAGES tools are currently undergoing pilot testing in locations in Africa, Southeast Asia, and South America and will be offered to other interested countries around the world.

ACKNOWLEDGMENTS: We acknowledge the support of Tim Frey, Supervisor of the Bio-Threat Awareness Systems Group in APL's Asymmetric Operations Sector. Joe Lombardo, an early and established authority in electronic disease surveillance, has provided helpful guidance to the SAGES program. Anna L. Buczak provided helpful comments on the manuscript.

REFERENCES

[1]Langmuir, A. D., "The Surveillance of Communicable Diseases of National Importance," *N. Engl. J. Med.* **268**, 182–192 (1963).

[2]Thacker, S. B., and Berkelman, R. L., "Public Health Surveillance in the United States," *Epidemiol. Rev.* **10**(1), 164–190 (1988).

[3]Thacker, S. B., "Historical Development," Chap. 1, *Principles of Public Health Surveillance*, S. M. Teutsch and R. E. Churchill (eds.), Oxford University Press, New York, pp. 3–17 (1994).

[4]U.S. Centers for Disease Control and Prevention, *Overview of Influenza Surveillance in the United States*, http://www.cdc.gov/flu/weekly/overview.htm (accessed 28 Feb 2014).

[5]Reingold, A., "If Syndromic Surveillance Is the Answer, What Is the Question?" *Biosecur. Bioterror.* **1**(2), 77–81 (2003).

[6]Council to Improve Foodborne Outbreak Response (CIFOR), *Guidelines for Foodborne Disease Outbreak Response*, Council of State and Territorial Epidemiologists, Atlanta (2009).

[7]Lombardo, J. S., and Ross, D., "Disease Surveillance: a Public Health Priority," Chap. 1, *Disease Surveillance: A Public Health Informatics Approach*, J. S. Lombardo and D. L. Buckeridge (eds.), John Wiley & Sons, Inc., Hoboken, pp. 1–39 (2007).

[8]Chretien, J. P., Burkom, H. S., Sedyaningsih, E. R., Larasati, R. P., Lescano, A. G., et al., "Syndromic Surveillance: Adapting Innovations to Developing Settings," *PLoS Med.* **5**(3), e72 (2008).

[9]Soto, G., Araujo-Castillo, R. V., Neyra, J., Fernandez, M., Leturia, C., Mundaca, C. C., and Blazes, D. L., "Challenges in the Implementation of an Electronic Surveillance System in a Resource-Limited Setting: Alerta, in Peru," *BMC Proc.* **2**(Suppl. 3), S4 (2008).

[10]The Centers for Disease Control and Prevention, "Status of State Electronic Disease Surveillance Systems—United States, 2007," *MMWR Morb. Mortal. Wkly. Rep.* **58**(29), 804–807 (2009).

[11]Jajosky, R. A., and Groseclose, S. L., "Evaluation of Reporting Timeliness of Public Health Surveillance Systems for Infectious Diseases," *BMC Public Health* **4**, 29 (2004).

[12]Lombardo, J., Burkom, H., Elbert, E., Magruder, S., Happel-Lewis, S., et al., "A System Overview of the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE II)," *J. Urban Health* **80**(2 Suppl 1), i32–i42 (2003).

[13]Binder, S., Levitt, A. M., Sacks, J. J., and Hughes, J. M., "Emerging Infectious Diseases: Public Health Issues for the 21st Century," *Science* **284**(5418), 1311–1313 (1999).

[14]Morens, D. M., Folkers, G. K., and Fauci, A. S., "The Challenge of Emerging and Re-emerging Infectious Diseases," *Nature* **430**, 242–249 (2004).

[15]The World Health Organization, *The World Health Report 2007 - a Safer Future: Global Public Health Security in the 21st Century*, WHO Press, Geneva (2007).

[16]Hollingsworth, T. D., Ferguson, N. M., and Anderson, R. M., "Frequent Travelers and Rate of Spread of Epidemics," *Emerg. Infect. Dis.* **13**(9), 1288–1294 (2007).

[17]World Health Organization, *WHO International Health Regulations (2005)*, 2nd ed., World Health Organization, Geneva (2008).

[18]Ashar, R., Lewis, S. L., Blazes, D. L., and Chretien, J. P., "Applying Information and Communications Technologies to Collect Health Data from Remote Settings: a Systematic Assessment of Current Technologies," *J. Biomed. Informatics* **43**(2), 332–341 (2010).

[19]Loschen, W., Coberly, J. S., Sniegoski, C., Holtry, R., Sikes, M. L., and Happel-Lewis, S., "Event Communication in a Regional Disease Surveillance System," *American Med. Informatics Assoc. Annual Proc.* 2007, 483–487 (2007).

[20]Feighner, B. H., Chretien, J. P., Murphy, S. P., Skora, J. F., Coberly, J. S., et al., "The Pandemic Influenza Policy Model: a Planning Tool for Military Public Health Officials," *Military Med.* **174**(6), 557–565 (2009).

[21]Happel-Lewis, S., Cutchis, P. N., Babin, S. M., and Burkom, H., "Simulated Release of Plague in Montgomery County, Maryland," *Johns Hopkins APL Tech. Dig.* **24**(4), 354–359 (2003).

[22]Buczak, A. L., Koshute, P. T., Babin, S. M., Feighner, B. H., and Lewis, S. H., "A Data-driven Epidemiological Prediction Method for Dengue Outbreaks Using Local and Remote Sensing Data," *BMC Med. Inform. Decis. Mak.* **12**, 124 (2012).

[23]Dean, A. G., Fagan, R. F., and Panter-Connah, B. J., "Computerizing Public Health Surveillance Systems," Chap. 11, *Principles and Practice of Public Health Surveillance*, S. M. Teutsch and R. E. Churchill (eds.), Oxford University Press, New York, pp. 200–217 (1994).

[24]Buehler, J. W., Hopkins, R. S., Overhage, J. M., Sosin, D. M., and Tong, V., "Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks," *MMWR Morb. Mortal. Wkly. Rep.* **53**(RR05), 1–11 (2004).

[25]Klaucke, D. N., Buehler, J. W., Thacker, S. B., Parrish, R. G., Trowbridge, F. L., Berkelman, R. L., and the Surveillance Coordination Group, "Guidelines for Evaluating Surveillance Systems," *MMWR Morb. Mortal. Wkly. Rep.* **37**(S-5), 1–18 (1988).

[26]Sosin, D. M., "Draft Framework for Evaluating Syndromic Surveillance Systems," *J. Urban Health* **80**(2 Suppl 1), i8–i13 (2003).

[27]Lombardo, J. S., Burkom, H., and Pavlin, J., "ESSENCE II and the Framework for Evaluating Syndromic Surveillance Systems," *MMWR Morb. Mortal. Wkly. Rep.* **53**(Suppl), 159–165 (2004).

The Authors

Brian H. Feighner is the Senior Public Health Physician on the SAGES team. **Timothy C. Campbell** is the technical lead in information technology for the SAGES project. **Aaron T. Katz** is Assistant Group Supervisor of the Bio-Threat Awareness Systems Group and the SAGES Project Manager. **Richard A. Wojcik** is the Bio-Threat Awareness Systems Group's Chief Engineer. **Jacqueline S. Coberly** is the technical lead in epidemiology for the SAGES project, and **Shraddha V. Patel** is the technical lead in education and training. **Wayne A. Loschen** leads APL's enterprise ESSENCE technical efforts. **Sheri H. Lewis** is the Global Health Surveillance Program Manager in the Asymmetric Operations Sector's Homeland Protection Program Management Office. For further information on the work reported here, contact Brian Feighner. His e-mail address is brian.feighner@jhuapl.edu.

The *Johns Hopkins APL Technical Digest* can be accessed electronically at **www.jhuapl.edu/techdigest**.