Methods for Information Sharing to Support Health Monitoring

Wayne A. Loschen

The practice of real-time surveillance of disease categories, sometimes called syndromic surveillance, is widespread at local, state, and national levels. The diseases identified by these systems, however, ignore health department jurisdictional boundaries. This creates situations where it is important for public health officials to be aware of conditions in other jurisdictions. There are currently two fundamental ways for systems to accomplish this: (i) share the underlying data or (ii) share information produced by the systems or their users. While many other efforts have concentrated on sharing data, the JHU/APL Center of Excellence in Public Health Informatics has been working on technology that allows information (rather than data) to be shared between multiple public health jurisdictions.

INTRODUCTION

The phrase "disease knows no boundaries" is a common one. When a disease outbreak begins spreading through a community, it will not stop at ZIP code, county, state, or national boundaries. This concept of a disease's ability to operate so freely is in contrast to the restrictions placed on the public health officials monitoring, investigating, and responding to these disease outbreaks. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) prohibits public health officials from capturing data containing any form of personal identifiers corresponding to patients outside of their jurisdictional authority.[1] However, there are situations where the only way to know whether a multijurisdictional outbreak is occurring is by collecting information on cases from multiple adjacent jurisdictions within a particular area. Because the law requires these boundaries, public health officials need to be able to monitor health situations outside of their boundaries in a HIPAA-compliant and legally acceptable way. This project was developed to facilitate that exchange of information between public health officials and specifically between public health users of disease surveillance systems.

Based on previously developed technology at APL for the Electronic Surveillance System for Early Notification of Community-based Epidemics (ESSENCE) disease surveillance system, the Information Exchange project will allow public health officials to create and exchange structured computer-readable messages about current disease outbreaks in their communities.[2] This information will be shared through web services to local applications and web pages containing maps, graphs, and the messages themselves. The exchange of information will allow better investigation into potential local threats and a more complete situational awareness of disease in the region; in addition, it will facilitate a community of individuals working closer together to protect the public's health.

DATA VERSUS INFORMATION

There are two fundamental ways for disease surveillance systems to communicate about disease outbreaks occurring between multiple jurisdictions: sharing data and sharing information. In this context, information and data are defined by what is being shared. In Fig. 1, data are defined as raw, cleaned, or aggregated data. Information is defined as univariate or multivariate analysis and epidemiological interpretations. Figure 2 shows examples of each category. Raw data are shown as emergency room visit records. These contain duplicate entries and abbreviations such as "SoB" instead of shortness of breath. The cleansed data example removes the duplicate record and expands the abbreviation. The aggregate data example gives the total number of records of emergency room respiratory visits and over-the-counter respiratory products sold. Univariate analysis is shown as a single detection result of a single data stream. Multivariate analysis is shown as the result of a fusion detector that has been alerted because of an increase in three of five different data streams. The final example is of epidemiological interpretations. It shows a written summary of the respiratory situation based on the epidemiologist's investigation of the situation. In Fig. 2, the scale from data to information is split between preanalysis and postanalysis, with information being the result of analyzing data. Both data and information can be shared between systems, but each has its advantages.

Disease surveillance systems are built to collect data, analyze it, and provide users access to both the data and the information created from that data. If data sharing could occur, systems are already in place that provide users with amazing levels of detail and specificity about the current health situation. Most disease surveillance systems include automated detection algorithms that can provide very specific results if given specific data elements. There are even algorithms that analyze individual patients for abnormal health behaviors that could pose public health threats.

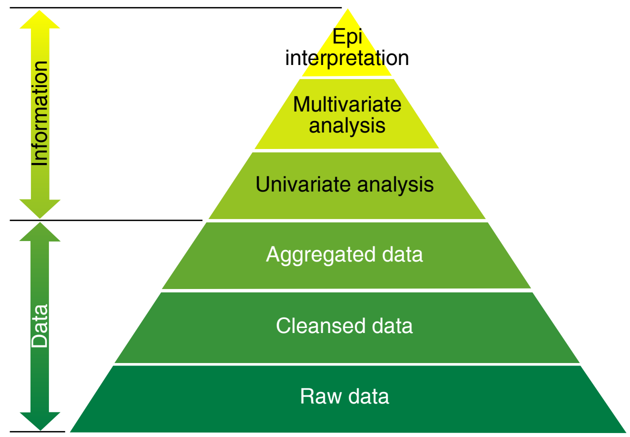

Figure 1. Definition of data versus information in the disease surveillance context. Epi, epidemiologist.

However, sharing data can be difficult, especially outside of the originating jurisdiction. It is expensive and time-consuming to create the legal agreements necessary to share data between different political jurisdictions, and in some cases it may be impossible. It can also be difficult to understand data from other jurisdictions. Although it is possible for anyone to incorrectly interpret the data they are viewing, local public health officials are in the best position to understand the nuances of data from their own community. A situation that might look like an outbreak to an outside observer might in fact be easily explained away by someone with local knowledge of a hospital closing, school function, or sporting event that an outsider would not know.

Sharing information instead of data provides some advantages. First, the legal issues are much easier to deal with. Sharing opinions about situations is less restrictive and requires fewer or no legal agreements. Second, because local system users are knowledgeable about their data and community, the result of their analysis provides an outside jurisdiction with the best understanding of what is or is not an outbreak. Users also provide their feedback in short summaries of the situation

Epi interpretation	Epi: "Respiratory outbreak we are currently investigating"
Multivariate analysis	Fusion detector: Respiratory has a Red Alert across 3 of 5 sources
Univariate analysis	Detector: ER Respiratory visits are four times the normal rate
Aggregated data	247 ER Respiratory visits, 1647 OTC Respiratory products sold
Cleansed data	John Doe, 26, M, Sore Throat Jane Doe, 20, F, Shortness of Breath . . .
Raw data	John Doe, 26, M, Sore Throat John Doe, 26, M, Sore Throat Jane Doe, 1987, Female, SoB . . .

Figure 2. Examples of data and information in a disease surveillance context. ER, emergency room; OTC, over the counter.

so that outside users can quickly review what is going on in another jurisdiction without having to do a complete analysis. The last major benefit is the increased sense of community that sharing information can provide. Users become familiar with other users and tend to collaborate and share additional information such as techniques, tools, and local policies that may be useful to others.

The disadvantages of sharing information include the loss of specificity that local algorithms will incur because they are not provided with patient-level information. The inputs to these algorithms will determine what they can do. If given information about a possible respiratory outbreak in a neighboring jurisdiction, they may be able to increase sensitivity in local respiratory detectors but will not be provided with exact numbers or locations to use. Although this may be enough to prompt a public health official to investigate a local event more closely, the official will be unable to further investigate the cases that caused the alert in the neighboring jurisdictions. If data were provided instead, the user could query for specific terms of interest.

CURRENT PRACTICE

Disease surveillance across multiple jurisdictions occurs today. ESSENCE has been developed to allow multiple jurisdictions to operate in the same system. Systems can span multiple health districts within a single state, or even multiple states, as shown in the National Capital Region Syndromic Surveillance Network (NCR-SSN).[2] These systems collect patient-level data and are considered the best way to detect and monitor outbreak situations. However, within these multijurisdictional systems, access control is in place to provide users only the data from their jurisdiction or aggregate data from other jurisdictions. This limits the data sharing that occurs between users but allows the underlying algorithms access to all of the data. In addition, all systems require legal agreements to be created between data providers and the system hosts to allow health data to be transferred to the public health officials. These legal agreements can sometimes be costly and time-consuming to create. In the case of the NCR-SSN, the legal agreements required years to complete because of the complexity of data being shared across state boundaries. If the jurisdictions already have a good working relationship, the legal agreements may be easier to create but will still require some amount of cost and time.

In addition to sharing data between users in systems that collect data across multiple jurisdictions, for the past 2 years ESSENCE systems have also provided users with an event communications component (ECC). This component provides users with a forum to discuss current abnormalities that they have seen and comment on their possibility of being a disease outbreak. Figure 3 shows a screen shot of the ECC user interface. It contains events that users have created to describe possible outbreaks occurring in the system. In this fashion, even if particular users cannot see detailed information from neighboring jurisdictions, they can at least discuss the interpretations of what is happening with a user that does have the rights to see the data.

The ECC provides users with a forum, but it also requires structured fields to be filled out with computer-readable selections. For example, if a user creates an event, they are prompted to fill out information about the date ranges, geographic regions, data sources, syndromes, and age groups in which the signal was found. When creating events or comments about events, they are also required to describe their level of concern by choosing one of five categories: informational only, not concerned, monitoring the situation, investigating the situation, and responding to an outbreak. This meta-information is then stored with the free-text message, is viewable by all users, and is usable by the system. An example of an event might include the message, "Strange increase in

Event List											
Details	Category	E-ID	Start Date	End Date	Rank	Comments	Status	Created By	Created On	Last Modified	**Edit**
Event Info	GI	22	08Jun05	08Jun05	High	2	Open	User Name Removed	01Jun06 03:00 PM EDT	01Jun06 03:23 PM EDT	
⊞ **Event Description:**	We have got total of 24 cases for today with GI. In addition to the RUQ abdominal pain we have got complaints of vomitus yellow, yellow eye color and yellow skin. Still high number of cases in zipcode 20001. Pts attend or are affiliated with Howard University.										
Event Info	Rash	21	08Jun05	08Jun05	Guarded	1	Open	User Name Removed	01Jun06 02:36 PM EDT	01Jun06 02:36 PM EDT	
⊞ **Event Description:**	Two boys 0-4 different zips, no info on discharge disp										
Event Info	GI	20	08Jun05	08Jun05	High	12	Open	User Name Removed	01Jun06 02:10 PM EDT	01Jun06 03:23 PM EDT	
⊞ **Event Description:**	Asking health districts if they have any lab info back from the investigation. Looks to us like it might be Hepatitis and we're looking for confirmation......										
Event Info	GI	19	07Jun05	07Jun05	Low	4	Open	User Name Removed	31May06 03:27 PM EDT	01Jun06 02:39 PM EDT	
⊞ **Event Description:**	Prince William has number higher than any in last 3 mos.										
									Simulated Data		

Figure 3. ESSENCE event list page from a simulated site. GI, gastrointestinal; RUQ, right upper quadrant.

respiratory; this might be the start of flu season." The same message may have meta-information that defines the region as Howard County, Maryland, the syndrome as respiratory, the data source as emergency room visits, the age group as 18–44 years old, the date as 1 October 2007, and the concern level as monitoring the situation. Each of the meta-information values is chosen from a list of available choices that the system understands. This allows the system to filter messages by specific, known meta-information fields. By using the meta-information, the system can create additional views that combine the results of mathematical detection algorithms and user concern levels on the same streams of data. An example of this combination of mathematical alerts and user-created concerns is shown in Fig. 4, which shows a screen shot of the summary alert list page found in ESSENCE. Each asterisk in the center of the page represents a day. The asterisks running along the top of a cell are detection results, and the asterisks running along the bottom are peer concerns. They are color-coded by the level of abnormality in the case of mathematical detection algorithm asterisks and the level of concern in the case of the user-defined event asterisks. With each cell representing a syndrome and geographic region, users viewing this page can quickly see patterns across regions and syndromes. They have the ability to also confirm or question mathematical results with peer concerns that are shown in the same cell. This combination of mathematical and epidemiological significance would not be possible without having collected the events with structured, computer-readable meta-information. It also shows that there are secondary uses of the information being shared beyond only visualizing the messages.

Although the ECC has shown researchers and users the value of structured, computer-readable information sharing, it only works within a single ESSENCE installation. Sharing this type of information between completely separate systems has not been done using the ECC. Sharing some types of data and information between different users and systems, however, has been accomplished by using other technologies and tools. These systems normally fall into one of two categories: sharing structured data or sharing unstructured information.

Systems that share structured data across jurisdictions include many disease surveillance systems, such as ESSENCE and BioSense [which is from the Centers for Disease Control and Prevention (CDC)].[3] These systems collect information from numerous hospitals and other data providers across many jurisdictions and provide aspects of those data to users who can be from other jurisdictions. Disease surveillance systems are designed to provide early detection and situational awareness of disease outbreaks in communities to public health officials. Regional Health Information Organizations (RHIOs) are also organizations that provide technology and support to share data across jurisdictions.[4] More clinical in nature, RHIOs are interested in sharing raw medical data between different stakeholders in the organization. Both systems, disease surveillance systems or RHIOs, transmit raw data between jurisdictions for use. There are other systems, such as the International Society for Disease Surveillance's DiSTRIBuTE project, that share aggregated data between sites.[5] Each jurisdiction submits aggregated counts, and selected graphs are created from the overall collection of data to be shared back out with their users. Although this system does not share raw data, it is still sharing data instead of information.

There are also systems that share unstructured information. The Program for Monitoring Emerging Diseases (ProMED-mail) is an example of a disease-reporting system that uses e-mail to distribute moderated reports.[6] These provide interpretations of outbreaks in free text. They are also moderated, which provides a level of quality assurance but also delays the delivery of the message and requires effort on the part of a moderating staff. The Health Alert Network has been developed by CDC with the National Association of County and City Health Officials, the Association of State and Territorial Health Officials, and other organizations to allow information to be shared among public health officials in the event of a health threat.[7] Again, this system allows unstructured text messages to be communicated between jurisdictions.

There is one other system that has been developed that comes close to collecting structured information about health outbreaks. The CDC-developed Epidemic

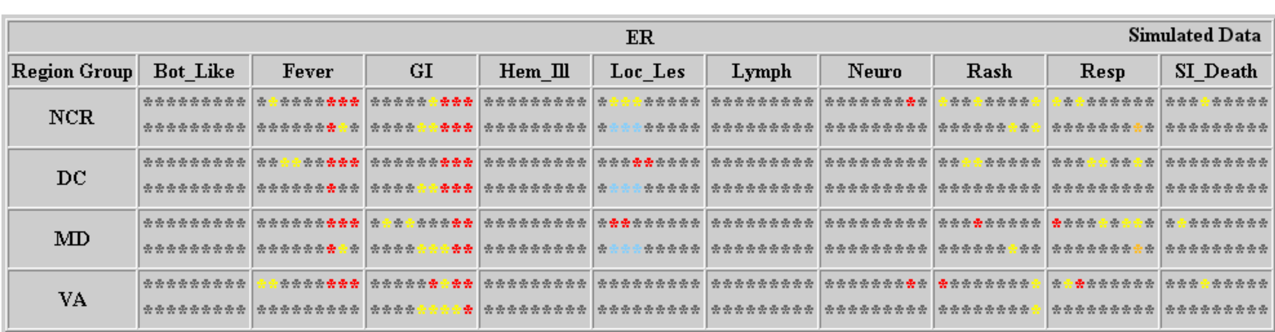

Figure 4. ESSENCE summary alert list page from a simulated site.

Information Exchange (Epi-X) is a program that shares moderated public health notices between public health officials.[8] Unlike ProMED-mail, the form to create a message requires some meta-information to be collected about the content of the message. Another difference is that Epi-X is limited to a more public health-oriented user base, and access is controlled. This controlled access, along with the lack of a computer-readable way for outside systems to digest the information, limits the use of the Epi-X tool for enhancing disease surveillance situational awareness to all users.

INFORMATION SHARING

With knowledge of the capabilities and limitations of current systems, researchers at APL began a project to develop a structured information-sharing system. The first major step in this project occurred during the 2007 National Football League Super Bowl. Super Bowl XLI took place at Dolphin Stadium in Miami, Florida. The teams that competed were the Indianapolis Colts and the Chicago Bears. All three of these locations, Miami, Indianapolis, and Chicago, are monitored by public health officials using ESSENCE systems. Four days before the Super Bowl, a conference call between APL researchers and public health officials in Miami, Indianapolis, the state of Indiana, and Cook County, Illinois, took place. During that call, the public health officials agreed to share information about their communities' health for the days before and a period of 2 weeks after the Super Bowl. The Miami version of ESSENCE was also modified to allow their users to watch for outbreaks in patients who lived in Indiana or Illinois. Miami public health officials already produced a daily summary report as part of their surveillance system. They agreed to share that report with officials from Indiana, Indianapolis, and Cook County who, in turn, agreed to inform the other jurisdictions if an outbreak was occurring. This summary report did not include details of any individuals but instead was a collection of epidemiological interpretations of what was going on with the health of their community. During this period of time, no outbreaks were reported.

Although this exercise did not include a structured message format or web-based site for collecting, distributing, and analyzing the messages, it did show the willingness of public health officials to share higher-level information between jurisdictions for weeks without having an ongoing outbreak already occurring. It also showed the ability to do so without the extended process of legal agreements. Finally, this exercise allowed four different groups of public health officials to network, collaborate, and form a better community that will help them understand how others utilize disease surveillance systems and better serve them if a public health emergency does occur in the future.

Building on the exercise, initial versions of message structures and designs for initial prototypes for the information-exchange system were created. Certain requirements have remained true throughout the life of this project. These include:

- Structured information from many different jurisdictions must be able to be shared.
- The messages created must be easily accessible for both users and systems.
- The information exchange must be able to accept these messages from a web interface or from a web service.
- The information exchange must be accessible by users of any disease surveillance system.

To create a system that meets these requirements, certain technologies have been selected. Initial prototypes include an XML-based message format, web services for information transmission between systems, and a web application to provide an interface for users. The XML-based message format allows any disease surveillance system the ability to parse and understand the information messages. Web services allow other systems the ability to transmit and receive those messages in an easy and computer-readable way. The web application will provide users of disease surveillance systems that are not currently integrated with the Information Exchange project the ability to participate by creating, viewing, and analyzing messages.

The initial message format will consist of the following fields:

- Syndrome
- Geography: Location
- Geography: Spread
- Age Group
- Sex
- Concern Level
- Event/Encounter Dates
- Race
- Hispanic Origin
- Event Size
- Case Definition
- Response Actions Taken
- Requests for Information
- Free-Text Comments
- Attachments

The syndrome, geography, and date fields will be required for all messages. All other fields will be optional when creating a message. Because structured messages are important, each field, excluding the request for information, free-text comment, and attachment fields, will be choices from a predetermined list of items. Syndromes will be a selection of the CDC's published list of bioterrorism-associated syndromes.[9] Country, state, county, ZIP code, or latitude and longitude will define geographic locations that can be used in the message. The geographic spread will be defined as either localized or widespread. Age groups will be broken down into 5- or 10-year blocks. The concern level will be a scale that is currently used in the ECC of ESSENCE. These categories include no concern, monitoring, investigating, or responding. Although these categories were originally low, medium, elevated, and high, we determined that action-oriented words were better understood by users and created less confusion. Race and Hispanic origin will be populated with the recommendations from current federal guidelines. Event size will allow the user to type in an exact number or choose from a list including: 1s, 10s, 100s, 1,000s, 10,000+. The case definition will be a logical expression of subsyndromes. The list of ~200 ESSENCE subsyndromes will be used initially until CDC publishes a standardized list. Other JHU/APL Center of Excellence projects are currently helping to create this standardized list. Response actions taken will be a list of common actions that public health officials take in the event of an outbreak. This list is still being developed, and we expect additions to be made to it during the first phases of this project's use. The request for information field will give users the ability to ask questions of the user who posted the message, and the free-text comment section will allow the creator a place to post extra information that does not fit into a structured field. We also believe that the ability to complement a message with an attachment may be useful, so we have planned for that capability.

Once an XML-based message has been created, it can be transmitted into the information exchange through a web service. The current prototype of the system uses Apache's Axis 2 as the web service framework. By using a web service to transmit and receive data, other disease surveillance systems can easily utilize the information in their own ways. We envision system developers creating their own visualizations and using the information to inform their own detection algorithms.

For those users whose systems have not integrated with the Information Exchange, a web application will provide users with web pages that allow message creation, viewing, and analysis. Original plans for this web application were to build it by using portal technologies. The intention of this plan was to allow users visiting the website to create personalized screens that could show any number of possible visualizations. These visualizations, including tables, geographic information system (GIS) maps, time series plots, and pie graphs, could be arranged on a personal webpage by each user. This technology, similar to the myYahoo and iGoogle personalized portal sites, provides a very flexible interface to users of different intentions and skill levels.[10,11] Initial prototypes of the system were developed by using Apache's Jetspeed 2 framework. However, users deemed this framework very cumbersome during an initial trial. The framework also did not allow each user to create customizable pages for themselves without extensive changes to the framework. The second framework we looked at was the JBoss Portal framework. This framework did support user customization of personal pages but again was not as easy for users to operate as we had hoped. The results of our initial prototypes using these frameworks were negative enough that the project has decided to not include a personalized portal-like user interface as a requirement but instead to focus on creating a clean, easy-to-use interface based on the ECC that we have already developed for the ESSENCE system.

In addition to the current ECC interface, the web application for this project will also need to include analysis screens that allow the user to quickly attain a level of situational awareness about the health of the participants' communities. These analysis screens should include GIS maps that show users where messages indicate potential outbreaks. These maps should provide users with an ability to filter down by each structured field to see specific syndromes, age groups, or concern levels. The current prototype creates these maps by using ESRI's ArcIMS software package. In addition to maps, users should be able to see time series plots and pie graphs that show statistics about the types of messages that are being created. These graphs help the users see what syndromes are being most talked about and which are of the highest concern. Finally, the user interface should allow users to subscribe to a notification system that will allow the system to contact the user if new messages arrive that meet their defined criteria. This way, if messages are created that show a high level of concern, they can be e-mailed to users directly instead of relying on the user to visit the site many times a day.

FUTURE PLANS

Once the initial version of the Information Exchange is complete, a set of current disease surveillance systems and specific users of those systems will be recruited to perform a limited trial of the exchange. These sites will predominately be made up of ESSENCE locations, but we hope to recruit at least one non-ESSENCE site to participate. Initial discussions with at least one non-ESSENCE site have been positive, and we are optimistic that we can accomplish this goal. Once the trial has been completed, surveys will be given to each of the participants. In these surveys, we hope to better understand what the system provided that was useful, as well as how the system could be improved in the future. We expect specific recommendations on message structure, message choices, and website features to be among the feedback we receive from the surveys.

Another future interest would be in the integration of the Information Exchange with other systems. Health officials utilize many different applications daily to perform their jobs, and incorporating public health information into those applications could provide a benefit to those users. An example application might be an emergency management tool. By feeding information from the public health information exchange into the tool, decision makers could have up-to-the-minute interpretations of a public health situation. By publishing the completed message structure, the project may find many different applications into which public health officials are interested in feeding information or from which they would like to obtain information.

CONCLUSION

Based on the previous success of the ECC of ESSENCE, it is clear that users of disease surveillance systems can benefit from sharing their interpretations among peers. Having the ability to communicate with neighboring jurisdictions about potential outbreaks has also proven to be useful in the National Capital Region. The next logical step is to allow users of different disease surveillance systems across the country and possibly across the world to share information about potential outbreaks. The ability to share this information in a structured, computer-readable format will allow disease surveillance systems to incorporate these data in many useful ways secondary to just visualizing the messages. It also will provide users with a situational awareness of the diseases occurring while lowering the possibility of misinterpreting data from non-local users of data. Utilizing local knowledge that is shared among all users can provide public health officials with the best possible analysis of potential outbreaks. This shared, structured information can help support the public health officials who are working hard each day to protect us all.

ACKNOWLEDGMENTS: This article and the work on this project would not have been possible without the following individuals: Joe Lombardo, Carol Sniegoski, Jacqueline Coberly, Nathaniel Tabernero, Vivian Hung, Radha Kowtha, Alimelu Jonnagadla, and the rest of the ESSENCE team. This article was supported by Grant P01 CD000270 from the CDC. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the CDC.

REFERENCES

[1] 104th Congress, *Public Law 104-191: Health Insurance Portability and Accountability Act of 1996* (21 Aug 1996), http://aspe.hhs.gov/admnsimp/pl104191.htm (accessed 17 Dec 2007).

[2] Lombardo, J. S., "The ESSENCE II Disease Surveillance Test Bed for the National Capital Area," *Johns Hopkins APL Tech. Dig.* **24**(4), 327–334 (2003).

[3] Centers for Disease Control and Prevention, *PHIN: BioSense Extramural Projects*, http://www.cdc.gov/biosense/extramuralprojects.htm (accessed 19 Dec 2007).

[4] Wikipedia contributors, "Regional Health Information Organization," *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org/wiki/RHIO (accessed 17 Dec 2007).

[5] International Society for Disease Surveillance, *DiSTRIBuTE Project*, http://www.syndromic.org/projects/DiSTRIBuTE.htm (accessed 17 Dec 2007).

[6] Madoff, L. C., "ProMED-mail: An Early Warning System for Emerging Diseases," *Clin. Infect. Dis.* **39**(2), 227–232 (2004).

[7] Centers for Disease Control and Prevention, *The Health Alert Network (HAN)*, http://www.bt.cdc.gov/documentsapp/HAN/han.asp (accessed 17 Dec 2007).

[8] Centers for Disease Control and Prevention, *Epi-X: Epidemic Information Exchange*, http://www.cdc.gov/epix (accessed 17 Dec 2007).

[9] Centers for Disease Control and Prevention, *Syndrome Definitions for Diseases Associated with Critical Bioterrorism-associated Agents* (23 Oct 2003), http://www.bt.cdc.gov/surveillance/syndromedef/index.asp (accessed 17 Dec 2007).

[10] myYahoo! website, http://cm.my.yahoo.com (accessed 17 Dec 2007).

[11] iGoogle website, http://www.google.com/ig (accessed 17 Dec 2007).

The Author

Wayne A. Loschen holds an M.S. in computer science from The Johns Hopkins University and a B.S. in computer science from the University of Illinois. He has worked at APL since 1999 as a software engineer, database administrator, system administrator, GIS developer, and web developer, and he has extensive experience in the management and visualization of large databases of environmental and health-related data. His work on the ESSENCE disease surveillance project has involved the development of novel detection algorithms, user interfaces, and visualization techniques for effectively understanding complex, multifaceted, and multistream temporal and spatial data. Mr. Loschen's e-mail address is wayne.loschen@jhuapl.edu.

Wayne A. Loschen