The Practice of Informatics JAMIA

Application of Information Technology ■

Technical Description of RODS: A Real-time Public Health Surveillance System

Fu-Chiang Tsui, PhD, Jeremy U. Espino, MD, Virginia M. Dato, MD, MPH,
Per H. Gesteland, MD, MS, Judith Hutman, Michael M. Wagner, MD, PhD

Abstract This report describes the design and implementation of the Real-time Outbreak and Disease Surveillance (RODS) system, a computer-based public health surveillance system for early detection of disease outbreaks. Hospitals send RODS data from clinical encounters over virtual private networks and leased lines using the Health Level 7 (HL7) message protocol. The data are sent in real time. RODS automatically classifies the registration chief complaint from the visit into one of seven syndrome categories using Bayesian classifiers. It stores the data in a relational database, aggregates the data for analysis using data warehousing techniques, applies univariate and multivariate statistical detection algorithms to the data, and alerts users of when the algorithms identify anomalous patterns in the syndrome counts. RODS also has a Web-based user interface that supports temporal and spatial analyses. RODS processes sales of over-the-counter health care products in a similar manner but receives such data in batch mode on a daily basis. RODS was used during the 2002 Winter Olympics and currently operates in two states—Pennsylvania and Utah. It has been and continues to be a resource for implementing, evaluating, and applying new methods of public health surveillance.

■ **J Am Med Inform Assoc.** 2003;10:399–408. DOI 10.1197/jamia.M1345.

Covert, large-scale attacks using biological agents such as anthrax, plague, tularemia, and smallpox can lead to massive casualties unless quarantine, vaccination, and/or antibiotic treatments are instituted promptly.[1] History highlights the need for timely detection of these threats. In 1979, there was an accidental release of anthrax from a bioweapons plant in Sverdlovsk, Russia. Before the anthrax outbreak was recognized, at least six patients exhibited influenzalike symptoms and were dismissed by their physicians as not having any serious illness. Twenty-one individuals had already died by the time laboratories confirmed the presence of *Bacillus anthracis*.[2]

Unfortunately, conventional public health disease surveillance—which relies on physician and laboratory reporting and manual analysis of surveillance data—is ill equipped for timely detection of such threats.[3] The reportable disease system relies on health care professionals to recognize, diagnose, and report cases and suspected outbreaks to public health officials[4,5]; however, it is unlikely that without an event or alert to raise his or her index of suspicion, a physician will attribute the early symptoms and signs of disease in a bioattack victim appropriately and report the case.[6] A key limitation of the current system is that the lone physician is blind to the cases his or her colleagues in a nearby hospital are seeing—knowledge that might lead the physician to consider uncommon diseases more strongly in his or her diagnostic reasoning. Mandatory laboratory reporting[4] is also ill-equipped for early detection, because it takes time before tests are ordered and specimens are obtained, transported, processed, and resulted.

Sufficiently early detection of a biological attack may be accomplished through surveillance schemes that can detect infected individuals earlier in the disease process. For completeness, we note that biosensors are being developed (and deployed) that detect organisms in the air and that this type of detection, if feasible, occurs fundamentally much earlier, because the delay introduced by the incubation period

Affiliations of the authors: University of Pittsburgh, Pittsburgh, Pennsylvania (F-CT, JUE, VMD (formerly), MMW); University of Utah, Salt Lake City, Utah (PHG); Intermountain Health Care, Salt Lake City, Utah (PHG). Dr. Dato is currently employed by the Pennsylvania Department of Health.

This work was supported in part by grants GO8 LM06625-01 and T15 LM/DE07059 from the National Library of Medicine; contract no. 290-00-0009 from the Agency for Healthcare Research and Quality; and by Cooperative Agreement Number U90/CCU318753-02 from the Centers for Disease Control and Prevention (CDC). Its contents are solely the responsiblity of the authors and do not necessarily represent the official views of CDC. Hoah-Der Su, Zhen Liu, and Xiaoming Zeng wrote components of RODS. The authors are indebted to Kimberlee Barnhart, Mary Cleatus Szczepaniak, participating hospitals, interface and network engineers, ESRI Inc., Bill Davenhall, and Lori Shienvold.

Correspondence and reprints: Fu-Chiang Tsui, PhD, The RODS Laboratory, Center for Biomedical Informatics, Suite 8084 Forbes Tower, 200 Lothrop Street, Pittsburgh, PA 15261; e-mail: <tsui@cbmi.pitt.edu>.

Received for publication: 02/19/03; accepted for publication: 05/13/03.

of the disease is eliminated from the surveillance system.[7] However, such approaches face unsolved technical problems in the analysis of contaminated specimens (the norm in air sampling). Biosensors also need to be in the right place—on every person's lapel or every street corner and hallway—to provide complete surveillance coverage.

Surveillance methods that can detect disease at an earlier stage are an important research direction for public health surveillance. These methods are generally referred to as *syndromic surveillance* because they have the goal of recognition of outbreaks based on the symptoms and signs of infection and even its effects on human behavior prior to first contact with the health care system.[8] Because the data used by syndromic surveillance systems cannot be used to establish a specific diagnosis in any particular *individual*, syndromic surveillance systems must be designed to detect signature *patterns of disease in a population* to achieve sufficient specificity. For example, it would be absurd to use only the symptom of fever to attempt to establish a working diagnosis of inhalational anthrax in an individual, but it would be very reasonable to establish a working diagnosis of anthrax release in a community if we were to observe a pattern of 1,000 individuals with fever distributed in a linear streak across an urban region consistent with the prevailing wind direction two days earlier. It would be beyond reasonable and, in fact, imperative to establish a working diagnosis of public health emergency if presented with such information.

One recent example of a form of syndromic surveillance is *drop-in surveillance*—the stationing of public health workers in emergency departments (EDs) and special clinics during high-profile events such as the Super Bowl to capture data on patients presenting with symptoms potentially indicative of bioterrorism. The major disadvantage of this approach is the cost of round-the-clock staffing for manual data collection.

A less expensive approach—and the one taken in the Real-time Outbreak and Disease Surveillance (RODS) system—is detection based on data collected routinely for other purposes. Examples of such data include absenteeism data, sales of over-the-counter (OTC) health care products, and chief complaints from EDs.[9] The expenses of manual data collection are avoided; however, the data obtained typically are noisy approximations of what could be obtained by direct interviewing of the patient (in the case of individual level data). Both approaches may play complementary roles with current methods of public health surveillance[10–12] by assisting the physician and public health official with a continuously updated picture of the "health status" of a population.[13,14]

A focus of our research has been syndromic surveillance from free-text chief complaints routinely collected by triage nurses in EDs and acute care clinics during patient registration. We have deployed this type of surveillance at the 2002 Winter Olympics and in the States of Pennsylvania and Utah. We described a previous version of the RODS system,[12] but the system has undergone considerable subsequent development both architecturally and functionally. This report provides a detailed description of the current version of RODS, an example of a computer-based public health surveillance system that adheres to the National Electronic Disease Surveillance System (NEDSS) specifications of the Centers for Disease Control and Prevention (CDC).[15,16]

Background

Public Health Surveillance

The role of public health surveillance is to collect, analyze, and interpret data about biological agents, diseases, risk factors, and other health events and to provide timely dissemination of collected information to decision makers.[17] Conventionally, public health surveillance relies on manual operations and off-line analysis.

Syndromic Surveillance

Existing syndromic surveillance systems include the CDC's drop-in surveillance systems,[8] Early Notification of Community-based Epidemics (ESSENCE),[10,18] the Lightweight Epidemiology Advanced Detection and Emergency Response System (LEADERS),[19] the Rapid Syndrome Validation Project (RSVP),[20] and the eight systems discussed by Lober et al.[11]

Lober et al. summarized desirable characteristics of syndromic surveillance systems and analyzed the extent to which systems that were in existence in 2001 had those characteristics.[11] A limitation of most systems (e.g., ESSENCE,[10] Children's Hospital in Boston,[11] University of Washington[11]) was batch transfer of data, which may delay detection by as long as the time interval (periodicity) between batch transfers. For example, a surveillance system with daily batch transfer may delay by one day the detection of an outbreak.

Some systems required manual data input (e.g., CDC's drop-in surveillance systems, RSVP,[20] and LEADERS[19]), which is labor-intensive and, in the worst case, requires round-the-clock staffing. Manual data input is not a feasible mid- or long-term solution even if the approach is to add items to existing encounter forms (where the items still may be ignored by busy clinicians).

A third limitation for existing surveillance systems is that the systems may not exploit existing standards or communication protocols like Heath Level 7 (HL7) even when they are available.

The data type most commonly used among surveillance systems is symptoms or diagnoses of patients from ED and/or physician office visits. Other types of data identified in that study include emergency call center and nurse advice lines. Other types of data being used include sales of over-the-counter health care products, prescriptions, telephone call volumes to health care providers and drug stores, and absenteeism. We have conducted studies demonstrating that the free-text chief complaint data that we use correlate with outbreaks.[21,22]

Design Objectives

The overall design objective for RODS is similar to that of an early warning system for missile defense; namely, to collect whatever data are required to achieve early detection from as wide an area as necessary and to analyze the data in a way that they can be used effectively by decision makers. It is required that this analysis be done in close to real time. This design objective is complex and difficult to operationalize because of the large number of organisms and the even larger number of possible routes of dissemination all requiring potentially different types of data for their detection, different algorithms, and different time urgencies. For this reason, our focus since beginning the project in 1999 has been on the specific problem of detecting a large-scale outbreak due to an outdoor (outside buildings) aerosol release of anthrax. Additional design objectives were adherence to NEDSS standards to ensure future interoperability with other types of public health surveillance systems, scalability, and that the system could not rely on manual data entry, except when it was done in a focused way in response to the system's own analysis of passively collected data.

Technical Description

This report describes RODS 1.5, which was completely rewritten as a Java 2 Enterprise Edition (J2EE) application since the previous publication describing it. RODS 1.5 is multidata type enabled, which means that any time series data can be incorporated into the databases and user interfaces. The deployed RODS system currently displays and analyzes health care delivery site registrations and separately monitors sales of OTC health care products.

Overview

RODS uses clinical data that are already being collected by health care providers and systems during the registration process. When a patient arrives at an ED (or an InstaCare in Utah), the registration clerk or triage nurse elicits the patient's reason for visit (i.e., the *chief complaint*), age, gender, home zip code, and other data and enter the data in a registration computer. The registration computer then generates an HL7 ADT (admission, discharge, and transfer) message and transmits it to the health system's HL7 message router (also called an *integration engine*). There usually is only one message router per health system even if there are many hospitals and facilities. These processes are all routine existing business activities and do not need to be created de novo for public health surveillance.

Figure 1 shows the flow of clinical data to and within RODS. The hospital's HL7 message router, upon receipt of an HL7 message from a registration computer, deletes identifiable information from the message and then transmits it to RODS over a secure virtual private network (VPN), or a leased line, or both (during the 2002 Winter Olympics we utilized both types of connections to each facility for fault tolerance). The RODS HL7 listener maintains the connection with the health system's message router and parses the HL7 message as described in more detail below. It then passes the chief complaint portion of the message to a Bayesian text classifier that assigns each free-text chief complaint to one of seven syndromic categories (or to an eighth category, *other*). The database stores the category data, which then are used by applications such as detection algorithms and user interfaces.

Data about sales of OTC health care products are processed separately by the National Retail Data Monitor, which is discussed in detail in another article in this issue of *JAMIA*.[23] The processing was kept separate intentionally because, in the future, the servers for the National Retail Data Monitor may operate in different physical locations than RODS. The RODS user interfaces can and do display sales of OTC health care products as will be discussed, but other user interfaces can be connected to the National Retail Data Monitor as well.

Data Level

Data Sharing Agreements

Prior to September 2001, RODS received data only from hospitals associated with the UPMC Health System, and efforts to recruit other hospitals met with resistance. After the terrorist attacks (including anthrax) in the Fall of 2001, other hospitals agreed to participate. Although data in this project are de-identified, certain information such as the number of ED visits by zip code were considered proprietary information by some health systems. Health Insurance Portability and Accountability Act (HIPAA) concerns also were very prominent in the discussions. Data-sharing agreements were

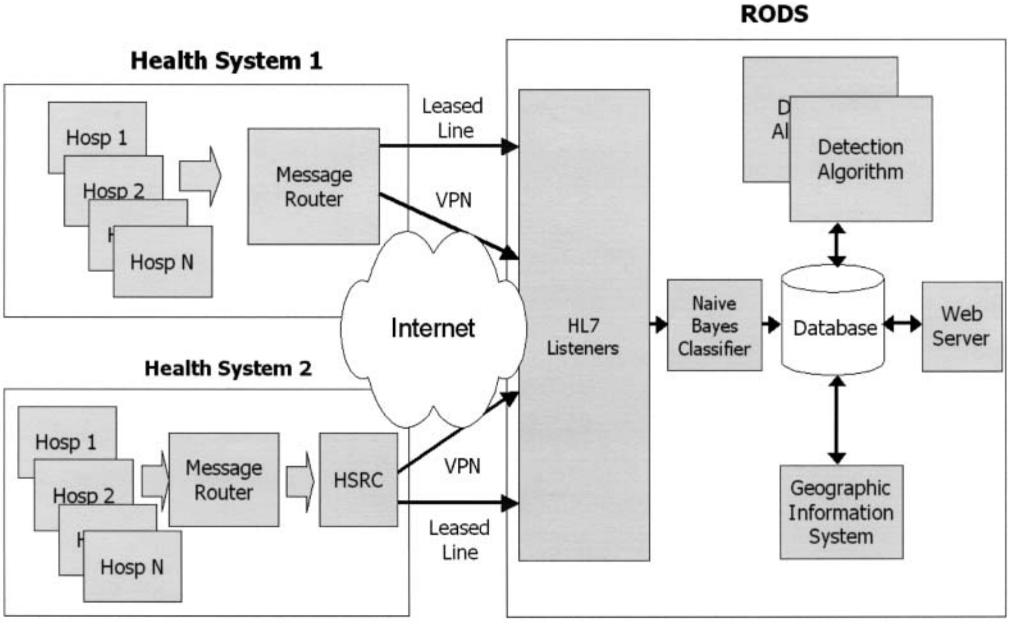


Figure 1. Data communication to RODS system from various health systems. (HSRC, health system resident component; VPN, virtual private network).

executed with every participating health system that addressed these concerns. As an additional precaution, all RODS project members meet annually with University of Pittsburgh council to review obligations and are required to sign an agreement every year stating that they understand the terms of the data-sharing agreements and agree to abide by the terms. RODS began as a research project at the University of Pittsburgh in 1999 and has functioned with IRB approvals since that time.

Data Types

Health care facilities send admission, discharge, and transfer (ADT) HL7 messages to RODS for patient visits in EDs and walk-in clinics. A minimal data set is sent, as shown in Figure 2, which qualifies as a HIPAA Limited Data Set.[24] Currently the data elements are age (without date of birth), gender, home zip code, and free-text chief complaint.

Data Transmission

The HL7 listener receives HL7 messages from the message routers located in each health system. The HL7 listener then passes the received HL7 message to the HL7 parser bean, an Enterprise JavaBean (EJB) in the RODS business logic tier. The HL7 parser bean uses regular expressions to parse the fields in an HL7 message. The HL7 parser bean then stores the parsed elements into a database through a managed database connection pool.

Although nearly all health systems utilize the HL7 messaging standard, the location of individual data elements in an HL7 message may differ from health system to health system. For example, some care providers' systems record free-text chief complaint in the DG1 segment instead of the PV2 segment of an HL7 message. To resolve this mapping problem, a configuration file written in eXtensible Markup Language (XML), a standard protocol often used to define hierarchical data elements, defines where each of the data elements can be found in the HL7 message. When an HL7 listener starts up, it reads the hospital-dependent configuration file and passes the configuration information to the parser bean.

We also use this configuration file to define the database table and field in which the HL7 parser bean should store each data element. This approach is useful because it allows the HL7 data to be stored to an external database. We anticipate that health departments with existing NEDSS or other public health surveillance databases may wish to use just this component of RODS for real-time collection of clinical data.

For hospitals that do not have HL7 message routers (two of approximately 60 in our experience to date), RODS accepts ED registration data files through either a secure Web-based data upload interface or a secure file transfer protocol. In general, these types of data transfers are technically trivial and for that reason are used by many groups but do not have the reliability of a HL7 connection (and have very undesirable time latencies).

Data Integrity

RODS checks the integrity of the data in the HL7 messages that it receives. This processing is necessary because hospital data flows may have undesirable characteristics such as duplicates. RODS identifies and deletes duplicates by using a database trigger that creates a composite primary key before inserting the data. RODS also filters out scheduling messages, which are identified by the fact that they have *future* admitted date and time.

RODS monitors all data feeds to ensure continuous connections with health systems. If RODS does not receive data for six hours, it sends an alert to the RODS administrator and the sending health system's administrator. Because the commercial message routers that hospitals use queue up HL7 messages when encountering networking or system problems, data integrity is preserved.

Database

RODS uses an Oracle8i database to store ED registration data. (Oracle, Redwood Shores, CA). To ensure fast response for an online query (e.g., the daily counts of respiratory syndrome in a county for the past six months), we developed a cache table scheme that pre-aggregates counts and refreshes them every 30 minutes.[25]

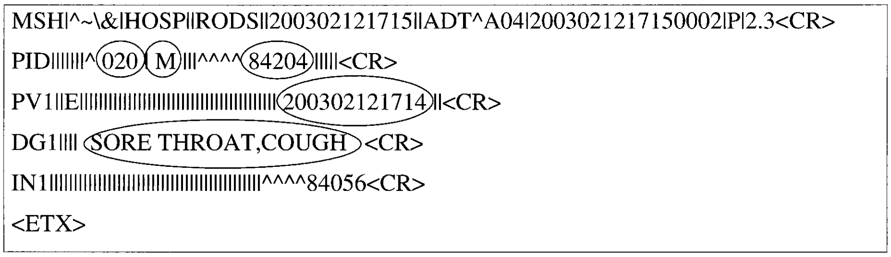
Network Level

The communications network between RODS and health care systems consists of virtual private networks (VPN) and leased lines. RODS uses multivendor site-to-site Internet Protocol Security (IPSEC) VPNs to receive HL7 messages. During the Winter Olympics, we exclusively used leased lines for the primary connection because of concerns about possible communications interruptions due to Internet traffic related to the games. The leased lines consisted of a redundant pair of 128k fractional T1 lines. After the Olympics, we returned to use of VPNs, and RODS has operated reliably using VPNs in both Utah and Pennsylvania. The leased-line modality is used only to connect the Siemens Medical Systems Data Center with RODS for the transmission of data from nine health systems that are hosted by Siemens.

System Hardware

For connectivity with the HL7 message routers, we utilize hardware-based routers. The VPN router is a Cisco PIX 501 and the leased-line routers are a pair of Cisco 2600s (Cisco Systems, Inc., San Jose, CA).

All of the RODS processes can be run on a single computer, but in our current implementation—serving Pennsylvania



```
MSH|^~\&|HOSP||RODS||200302121715||ADT^A04|20030212171500002|P|2.3<CR>
PID|||||||^020|M|||^^^^84204|||||<CR>
PV1||E||||||||||||||||||||||||||||||||||||||||||200302121714||<CR>
DG1|||| SORE THROAT,COUGH <CR>
IN1|||||||||||||||||||||||||||||||||||||||||^^^^84056<CR>
<ETX>
```

Figure 2. Sample HL7 admission, discharge, and transfer (ADT) message from an emergency department. The circled fields are age, gender, home zip code, admitted date and time, and free-text chief complaint, respectively.

and Utah as an application service provider—we use five dedicated servers: firewall, database, Web server, a geographic information system (GIS) server, and computation. The processes are written in Java code and can run on most platforms, but here we describe the specific platforms we use to indicate approximate sizing and processing requirements.

The database server is a Sun Microsystems Enterprise 250 configured with two Ultrasparc II 400Mhz processors, 2 gigabytes of RAM, and 36 gigabytes of mirrored hard drive space running an Oracle 8.1.7 (database) on Solaris 8 (Sun Microsystems, Inc., Santa Clara, CA).

The Web server is a Dell Poweredge 1550 configured with two 1Ghz Pentium III processors, 1 gigabyte of RAM, and 36 gigabytes of Redundant Arrays of Inexpensive Disk 5 (RAID-5) storage running Apache 1.3.24 (Web server), and Jboss 3.0 (described below in Fault Tolerance) on Redhat Linux 7.1 (Dell Computer Corporation, Round Rock, TX; Jboss Group, Atlanta, GA; Red Hat, Raleigh, NC).

The GIS server is a Dell Poweredge 350 configured with one 1Ghz Pentium III processor, 512 megabytes of RAM, and 18 gigabytes of storage running ArcIMS 4.0 (ESRI, Inc., Redlands, CA), an Internet-enabled geographic information system on Redhat Linux 7.3.

The computation server is a Penguin Computing server configured with dual Athlon MP 2400s, 1 gigabyte of RAM, and 750 gigabytes of RAID-5 storage running Oracle 9i on Redhat Linux 7.3.

Backup is performed nightly on all machines using a Sun StoreEdge L9 Tape Autoloader attached to the database server and Veritas Netbackup software (Veritas, Mountain View, CA).

Application Level

We developed RODS applications using the Java 2 Enterprise Edition Software Toolkit (J2EE SDK) from Sun Microsystems for cross-platform Java application development and deployment.[26]

We followed contemporary application programming practices—a multitiered application consisting of a client tier (custom applications such as HL7 listeners and detection algorithms), business logic tier, database tier, and Web tier.

Business logic such as the HL7 parser bean was implemented as Enterprise JavaBeans (EJBs). NEDSS specifies EJB as the standard for application logic. RODS uses Jboss, an open-source J2EE application server, to run all EJBs.[10]

The Web tier comprises the graphical user-interface to RODS and uses Java Server Pages (JSP), Java Servlets, and ArcIMS. The database tier was implemented in Oracle 8i.

Natural Language Processing

RODS uses a naive Bayesian classifier called *Complaint Coder* (CoCo) to classify free-text chief complaints into one of the following syndromic categories: constitutional, respiratory, gastrointestinal, neurological, botulinic, rash, hemorrhagic, and other. CoCo computes the probability of each category, conditioned on each word in a free-text chief complaint and assigns a patient to the category with the highest probability.[27] The probability distributions used by CoCo are learned from a manually created training set. CoCo can be retrained with local data, and it can be trained to detect a different set of syndromes than we currently use. CoCo runs as a local process on the RODS database server. CoCo was developed at the University of Pittsburgh and is available for free download at <http://health.pitt.edu/rods/sw>.

Detection Algorithms

Over the course of the project, RODS has used two detection algorithms. These algorithms have not been formally field tested because the emphasis of the project to date has been on developing the data collection infrastructure more than field testing of algorithms.

The Recursive-Least-Square (RLS) adaptive filter[28] currently runs every four hours, and alerts are sent to public health officials in Utah and Pennsylvania. RLS, a dynamic autoregressive linear model, computes an expected count for each syndrome category for seven counties in Utah and 16 counties in Pennsylvania as well as for the combined counts for each state. We use RLS because it has a minimal reliance on historical data for setting model parameters and a high sensitivity to rapid increases in a time series e.g., a sudden increase in daily counts. RLS triggers an alert when the current actual count exceeds the 95% confidence interval for the predicted count.

During the 2002 Olympics we also used the What's Strange About Recent Events (WSARE 1.0) algorithm.[29] WSARE performs a heuristic search over combinations of temporal and spatial features to detect anomalous densities of cases in space and time. Such features include all aspects of recent patient records, including syndromal categories, age, gender, and geographical information about patients. The criteria used in the past for sending a WSARE 1.0 alert was that there has been an increase in the number of patients with specific characteristics relative to the counts on the same day of the week during recent weeks and the p-value after careful adjustment for multiple testing for the increase was ≤ 0.05. Version 3.0 of WSARE, which will incorporate a Bayesian model for computing expected counts rather than using unadjusted historical counts currently, is under development.

Alert Notification

When an algorithm triggers an alert based on the above criteria, RODS sends e-mail and/or page alerts to its users. RODS uses an XML-based configuration file to define users' e-mail and pager addresses. The e-mail version of the alert includes a URL link to a graph of the time series that triggered the alarm with two comparison time series: total visits for the same time period and normalized counts.

User Interface

RODS has a password-protected, encrypted Web site at which users can review health care registration and sales of OTC health care products on epidemic plots and maps. When a user logs in, RODS will check the user's profile and will display data only for his or her health department's jurisdiction. The interface comprises three screens—Main, Epiplot, and Mapplot.

The *main screen* alternates views automatically among each of the available data sources (currently health care registrations and OTC products in Pennsylvania and Utah and OTC sales only for other states). The view alternates every two minutes as shown in Figure 3. The *clinic visits view* shows daily total visits and seven daily syndromes for the past week. The *OTC data view* shows daily sales for five product categories and the total, also for the past week. Users also can set the view to a specific county in a state. If the *normalize* control box is checked, the counts in the time series being displayed will be divided by (normalized by) the total daily sales of OTC health care products or ED visits for the region.

The *Epiplot* screen provides a general epidemic plotting capability. The user can simultaneously view a mixture of different syndromes and OTC product categories for any geographic region (state, county, or zip code), and for any time interval. The user also can retrieve case details as shown in Figure 4. The *Get Cases* button queries the database for the admission date, age, zip code, and chief complaint (verbatim, not classified into syndrome category) of all patients in the time interval and typically is used to examine an anomalous density (spike) of cases. The *Download Data* button will download data as a compressed comma separated file for further analyses.

The *Mapplot* screen is an interface to ArcIMS, an Internet-enabled GIS product developed by Environmental Systems Research Institute, Inc. Mapplot colors zip code regions to indicate the proportion of patients presenting with a particular syndrome. The GIS server also can overlay state boundaries, county boundaries, water bodies, hospital locations, landmarks, streets, and highways on the public health data as shown in Figure 5. Similar to Epiplot, Mapplot also can display case details for a user-selected zip code.

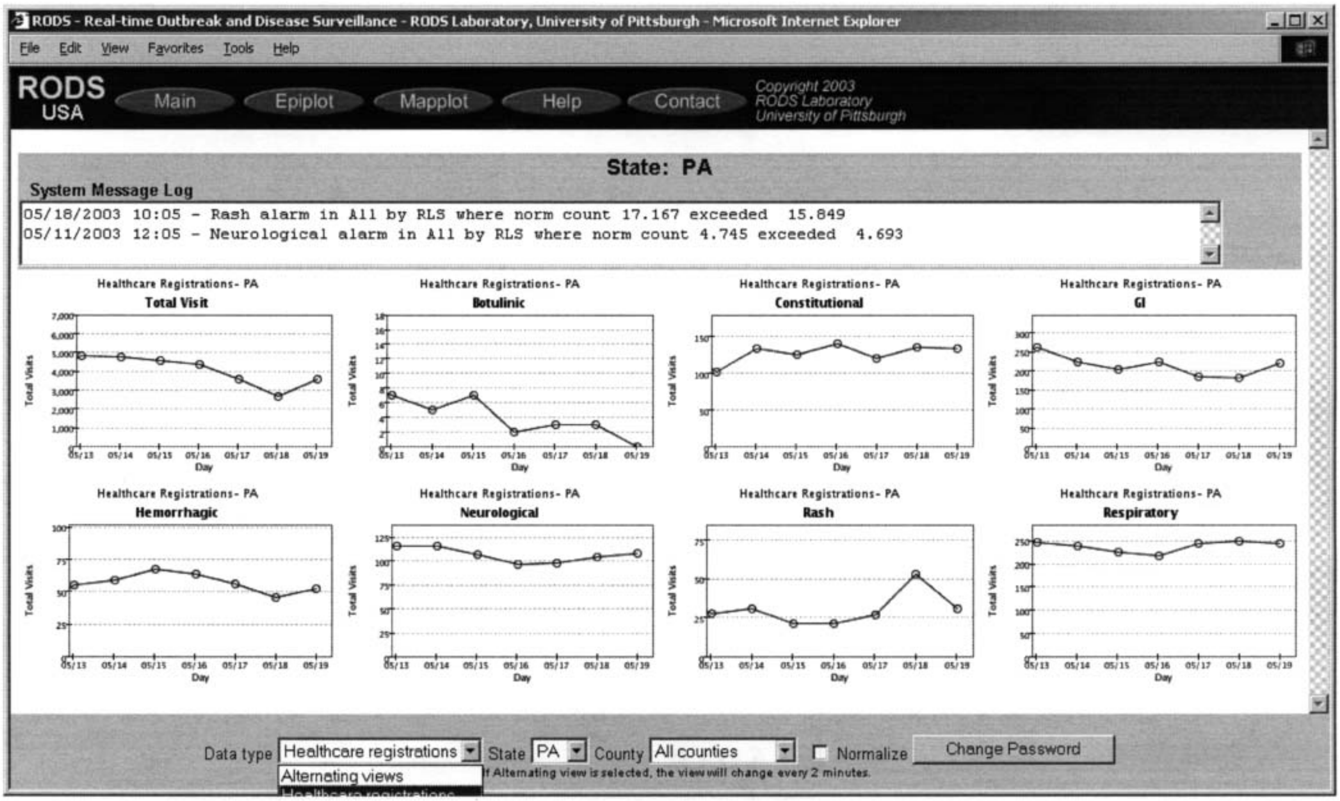
Fault Tolerance

RODS has been in operation for four years and, like most production systems, has acquired many fault-tolerant features. For example, at the software level, HL7 listeners continue to receive messages and temporarily store the messages when the database is off-line. A data manager program runs every ten minutes and, on finding such a cache, it loads the unstored messages to the database when the database is back on-line. In addition, the data manager program monitors and restarts HL7 listeners as necessary. The database uses "archive log" mode to log every transaction to ensure that the database can recover from a system failure.

The hardware architecture also is fault tolerant. All servers have dual power supplies and dual network cards. All hard drives use Redundant Arrays of Inexpensive Disk configurations. In addition to dual power supplies, all machines are connected to an uninterrupted power supply that is capable of sending an e-mail alert to the RODS administrator when the main power is down.

Health System Resident Component

An important component of RODS that currently is used only at the UPMC Health System in Pittsburgh is the Health System Resident Component (HSRC). The HSRC is located within the firewall of a health system and connects directly to the HL7 message router. The HSRC currently receives a diverse set of clinical data from the HL7 message router including culture results, radiology reports, and dictated


RODS - Real-time Outbreak and Disease Surveillance - RODS Laboratory, University of Pittsburgh - Microsoft Internet Explorer
File Edit View Favorites Tools Help
RODS USA
Main
Epiplot
Mapplot
Help
Contact
Copyright 2003 RODS Laboratory University of Pittsburgh
State: PA
System Message Log
Healthcare Registrations- PA
Total Visit
Botulinic
Constitutional
GI
Hemorrhagic
Neurological
Rash
Respiratory
Total Visits
Day
05/13 05/14 05/15 05/16 05/17 05/18 05/19
Data type
Alternating views
Change Password

```
05/18/2003 10:05 - Rash alarm in All by RLS where norm count 17.167 exceeded  15.849
05/11/2003 12:05 - Neurological alarm in All by RLS where norm count 4.745 exceeded  4.693
```

Data type: Healthcare registrations

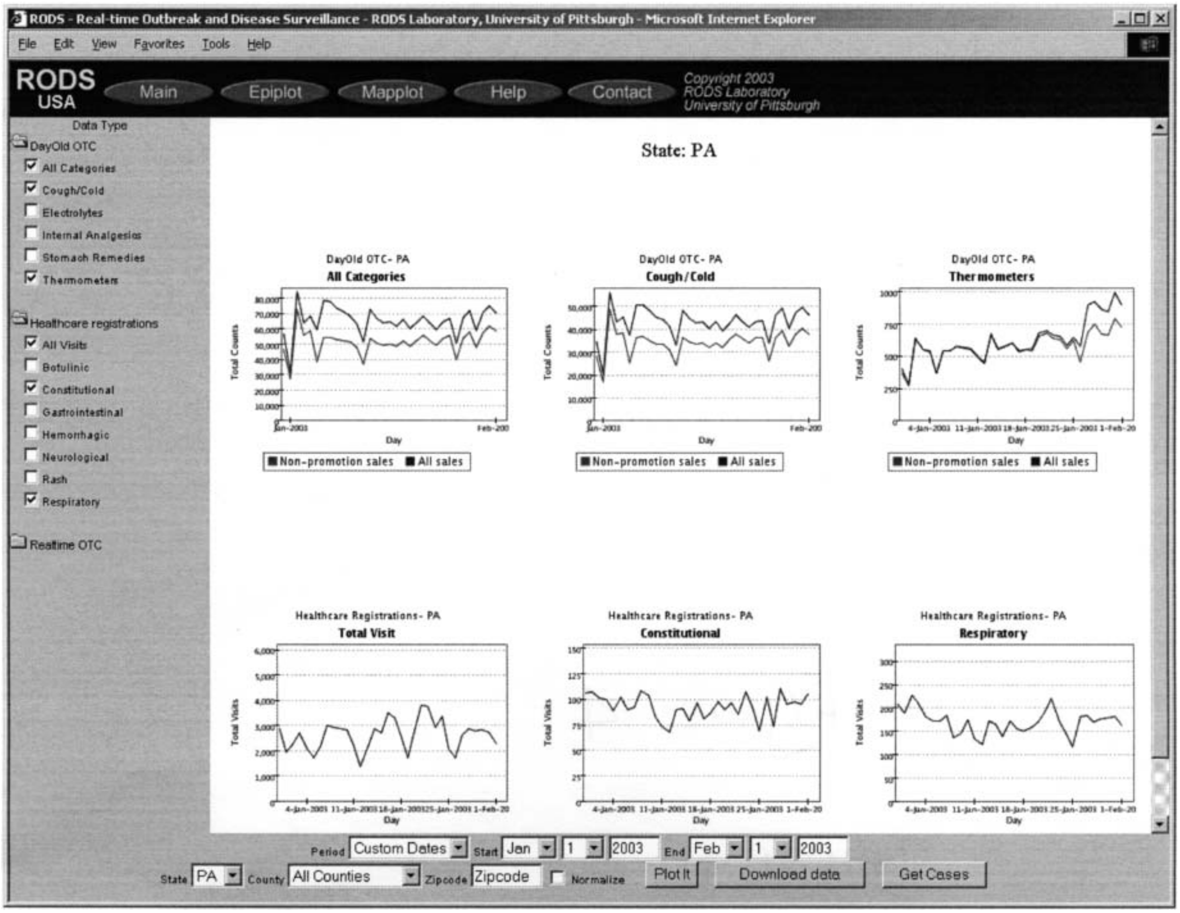
State: PA

County: All counties

☐ Normalize

If Alternating view is selected, the view will change every 2 minutes.

Figure 3. *Health care registrations view* in the Main screen of RODS. The Main screen alternates views every 2 minutes among data types available in the public health jurisdiction. The figure shows eight plots of health care registration data—total visits, botulinic, constitutional, gastrointestinal (GI), hemorrhagic, neurological, rash, and respiratory. After 2 minutes, over-the-counter data will be displayed. The Main screen can be used as a "situation room" display.


RODS - Real-time Outbreak and Disease Surveillance - RODS Laboratory, University of Pittsburgh - Microsoft Internet Explorer
File Edit View Favorites Tools Help
RODS USA
Main
Epiplot
Mapplot
Help
Contact
Copyright 2003 RODS Laboratory University of Pittsburgh
State: PA
DayOld OTC- PA
All Categories
Cough/Cold
Thermometers
Healthcare Registrations- PA
Total Visit
Constitutional
Respiratory
Total Counts
Total Visits
Day
Jan-2003
Feb-200
4-Jan-2003 11-Jan-2003 18-Jan-2003 25-Jan-2003 1-Feb-20
■ Non-promotion sales ■ All sales
Plot It
Download data
Get Cases

Data Type

DayOld OTC
- ☑ All Categories
- ☑ Cough/Cold
- □ Electrolytes
- □ Internal Analgesics
- □ Stomach Remedies
- ☑ Thermometers

Healthcare registrations
- ☑ All Visits
- □ Botulinic
- ☑ Constitutional
- □ Gastrointestinal
- □ Hemorrhagic
- □ Neurological
- □ Rash
- ☑ Respiratory

Realtime OTC

Period Custom Dates Start Jan 1 2003 End Feb 1 2003

State PA County All Counties Zipcode Zipcode □ Normalize

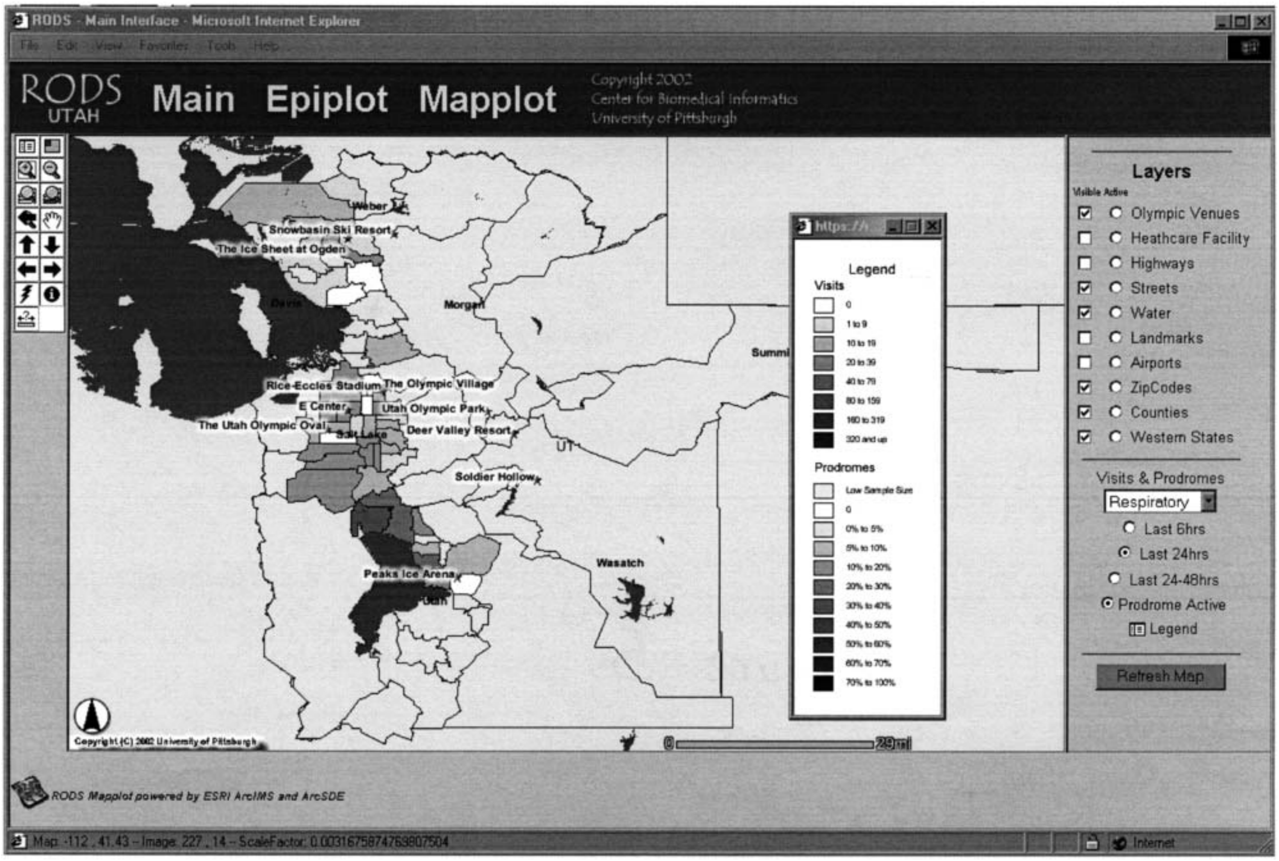
Figure 4. *Epiplot* screen of RODS. The six graphs are user-selected time-series plots of emergency department visits and sales of over-the-counter (OTC) products in Pennsylvania—OTC all categories, OTC cough/cold, OTC thermometers, clinic total visits, clinic constitutional, and clinic respiratory—between January 1, 2003, and February 1, 2003. Users can view a mixture of different syndromes and OTC product categories for any geographic region (state, county, or zip code), and for any time interval. Users can select types of data from the pick-list on the left of the screen. The *Download data* button retrieves raw count data for the selected graphs to a compressed comma-separated file. The *Get Cases* button shows a list of records containing chief complaint, age in decile, gender, and patient home zip code within the specified time interval and the geographic region. The lower, red line in the OTC plots represents nonpromoted sales.

emergency room notes. Its purpose is to provide additional public health surveillance functions that would not be possible if it were located outside of the firewall due to restrictions on the release of identifiable clinical data. The HSRC uses patient identifiers to link laboratory and radiology information to perform case detection. In the past, we have used HSRC to monitor for patients with both a gram-positive rod in a preliminary microbiology culture report and "mediastinal widening" in a radiology report. The HSRC is a case detector in a distributed outbreak detection system that is capable of achieving much higher specificity of patient diagnostic categorization through access to more information.

HSRC also removes identifiable information before transmitting data to the RODS system, a function provided by the health system's message router in other hospitals that connect to RODS.

The HSRC at UPMC Health System functions as an electronic laboratory reporting system, although the state and local health departments are not yet ready to receive real-time messaging from the system. Currently, it sends email alerts to the director of the laboratory and hospital infection control group about positive cultures for organisms that are required to be reported to public health in the state of Pennsylvania.[30] It also sends messages to hospital infection control when it detects organisms that cause nosocomial infections. These organisms include *Clostridium difficile*, methicillin-resistant *Staphylococcus aureus*, and vancomycin-resistant *Enterococcus*.

We have been able in HSRC to prototype one additional feature, which is a "look-back" function that facilitates very rapid outbreak investigations by providing access to electronic medical records to public health investigators as shown in Figure 6. This feature requires a token that can be passed to a hospital information system that can uniquely identify a patient, and the reason we have prototyped this feature in the HSRC and not in RODS is simply that HSRC runs within the firewall so an unencrypted token can be used. The look-back is accomplished as follows: when a public health user identifies an anonymous patient record of interest (e.g., one of 20 patients with diarrhea today from one zip code), HSRC calls the UPMC Health System Web-based electronic medical record system and passes it the patient identifier. UPMC Health System then requests the user to log in using the UPMC-issued password before providing access to the record



Layers

Visible Active

☑ ○ Olympic Venues
□ ○ Heathcare Facility
□ ○ Highways
☑ ○ Streets
☑ ○ Water
□ ○ Landmarks
□ ○ Airports
☑ ○ ZipCodes
☑ ○ Counties
☑ ○ Western States

Visits & Prodromes

Respiratory

○ Last 6hrs
● Last 24hrs
○ Last 24-48hrs
● Prodrome Active

Map: -112 , 41.43 – Image: 227 , 14 – ScaleFactor: 0.0031675874763807504

Figure 5. *Mapplot* screen of RODS shows spatial distribution of respiratory cases in part of Utah. Olympic venues are labeled.

directly from its own secure Web site. This approach is not intended to be implemented in HSRC, but rather in the RODS system outside of the firewall of a health system. It is intended to use encrypted identifiers that the health system would decrypt to retrieve the correct record. The HSRC could provide the encryption-decryption service or it could be provided by another data system in the hospital. We estimate that the prevalence of health systems that have Web-based results review in the United States is 30% to 50% and growing so that this approach could very quickly improve the efficiency of outbreak investigations.

Current Status

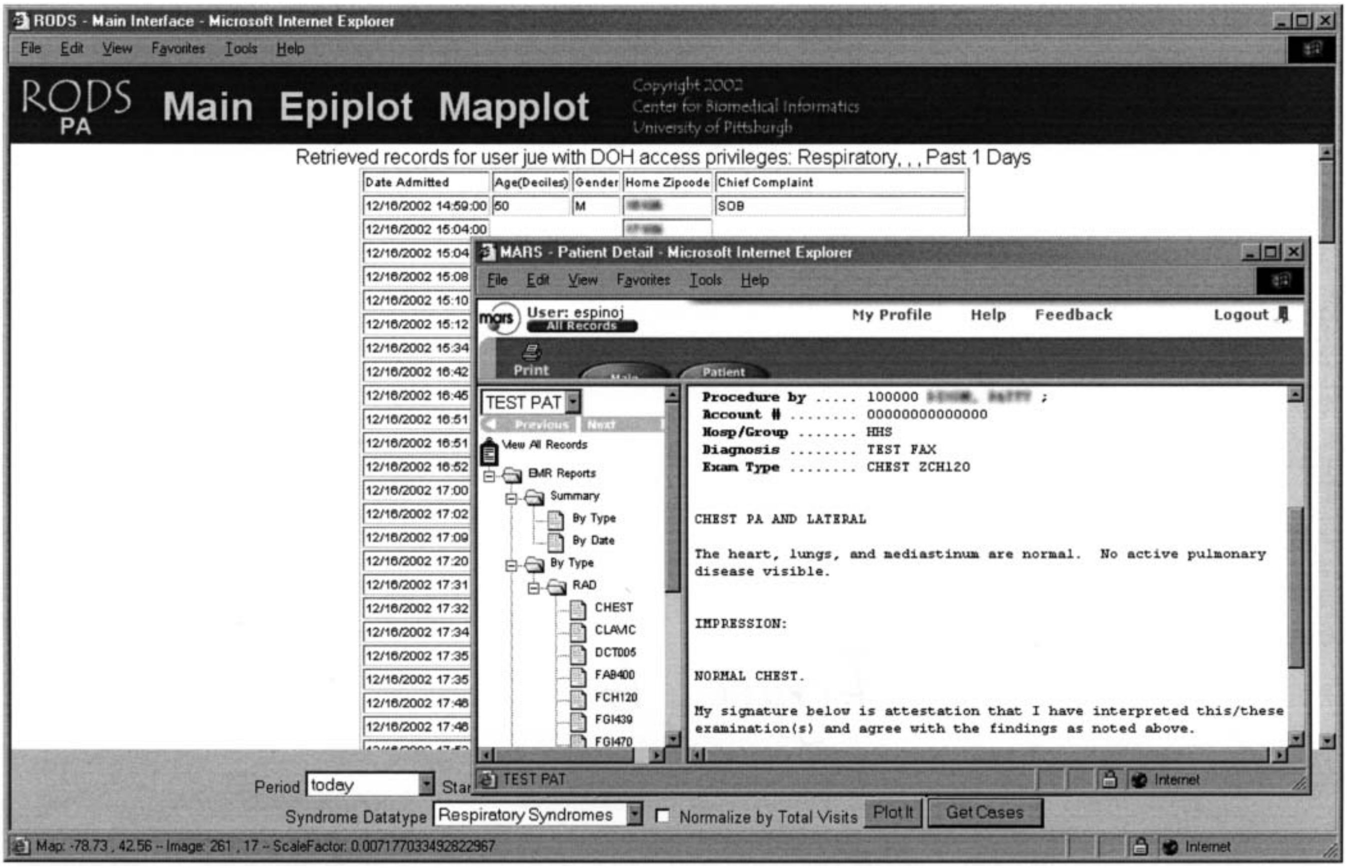
RODS has been in operation in Pennsylvania since 1999 and in Utah since January 2002. In Utah, RODS receives data from two health systems: Intermountain Heath Care, including nine EDs and 18 acute care facilities, and the University of Utah Health Sciences Center, with one ED.[24] Together, these facilities serve about 70% of the population of Utah. In Pennsylvania, RODS receives data from 20 health systems comprising 38 hospitals. Two health systems (each with one hospital) send plain text files to RODS on a daily basis. In Pennsylvania, RODS covers 80% of ED visits in Allegheny County (population, 1.3 M) where Pittsburgh is located; 50% of visits in the 13-county Metropolitan Medical Response Area centered on Pittsburgh (population 3.0 M); and more than 70% coverage of three other counties, including Dauphin County where Harrisburg, the capital of Pennsylvania, is located. The Commonwealth of Pennsylvania is funding a large project to connect the remaining hospitals in the Commonwealth with RODS over the next two years (approximately an additional 170 hospitals).

In December 2002, the RODS laboratory released version 1.1 of the RODS software to the public. The release includes all of the components necessary to deploy RODS for clinic visits surveillance. RODS is free for noncommercial use and can be downloaded at <http://www.health.pitt.edu/rods/sw/>. Although the software has been downloaded in excess of 170 times, we are aware of only a few successful efforts at deployment. These kinds of systems require network engineers, Oracle database administrators, and interface engineers, and very few health departments have access to that skills set.

For these reasons, we have moved to an application service provider model for dissemination in which we encourage state and local health departments to form coalitions to support shared services. We also have been fortunate to have sufficient grant funding from the Commonwealth of Pennsylvania to be able to support these services on an interim basis while sustainable funding models evolve.

Discussion

Our original design objectives for RODS were real-time collection of data with sufficient geographic coverage and sampling density to provide early syndromic warning of a large-scale aerosol release of anthrax. Although we have not achieved all of our initial design objectives, progress has been substantial. The research identified two types of data—free-text chief complaints and sales of OTC health care prod-


RODS - Main Interface - Microsoft Internet Explorer
File Edit View Favorites Tools Help
RODS PA
Main Epiplot Mapplot
Copyright 2002
Center for Biomedical Informatics
University of Pittsburgh
MARS - Patient Detail - Microsoft Internet Explorer
File Edit View Favorites Tools Help
mars
My Profile Help Feedback Logout
Print
Main
Patient
Previous Next
View All Records
EMR Reports
Summary
By Type
By Date
By Type
RAD
CHEST
CLAMC
DCT005
FAB400
FCH120
FGI439
FGI470
Internet
Plot It
Get Cases
Internet

Retrieved records for user jue with DOH access privileges: Respiratory, , , Past 1 Days

Date Admitted	Age(Deciles)	Gender	Home Zipcode	Chief Complaint
12/16/2002 14:59:00	50	M	[redacted]	SOB
12/16/2002 15:04:00			[redacted]	
12/16/2002 15:04				
12/16/2002 15:08				
12/16/2002 15:10				
12/16/2002 15:12				
12/16/2002 15:34				
12/16/2002 16:42				
12/16/2002 16:45				
12/16/2002 16:51				
12/16/2002 16:51				
12/16/2002 16:52				
12/16/2002 17:00				
12/16/2002 17:02				
12/16/2002 17:09				
12/16/2002 17:20				
12/16/2002 17:31				
12/16/2002 17:32				
12/16/2002 17:34				
12/16/2002 17:35				
12/16/2002 17:35				
12/16/2002 17:46				
12/16/2002 17:46				

User: espinoj
All Records

TEST PAT

```
Procedure by ..... 100000 [redacted] ;
Account # ........ 00000000000000
Hosp/Group ....... HHS
Diagnosis ........ TEST FAX
Exam Type ........ CHEST ZCH120


CHEST PA AND LATERAL

The heart, lungs, and mediastinum are normal.  No active pulmonary
disease visible.


IMPRESSION:


NORMAL CHEST.

My signature below is attestation that I have interpreted this/these
examination(s) and agree with the findings as noted above.
```

TEST PAT

Period today Star

Syndrome Datatype Respiratory Syndromes □ Normalize by Total Visits

Map: -78.73 , 42.56 – Image: 261 , 17 – ScaleFactor: 0.007177033492822967

Figure 6. Look-back function of RODS. The user has selected one patient to investigate using the screen that is in the background and partly hidden by overlap. RODS has logged the user into the results-review function of an electronic medical record and requested that patient's chart, which is shown on the screen in the foreground.

ucts—that can be obtained in real time or near real time at sampling levels of 70% or higher for most of the United States. These results were obtained through large-scale deployments of RODS in Pennsylvania and Utah and through building the National Retail Data Monitor described in the accompanying article in this issue of *JAMIA*. The deployments also provided insights about organizational and technical success factors that would inform an effort to scale the project nationally.

The project established the importance of HL7 message routers (also known as integration engines) for public health surveillance. HL7 message routers are a mature, highly prevalent technology in health care. We demonstrated that free-text triage chief complaints can be obtained in real time from most U.S. hospitals through message routers and that these data represent early syndromal information about disease. Many other clinical data of value to public health are transmitted using the HL7 standard (e.g., orders for diagnostic tests, especially microbiology tests, reports of chest radiographs, medications, and test results) and can be integrated into RODS or other surveillance systems capable of receiving HL7 messages.

As a result of our efforts to disseminate this technology by giving it away, we have learned that most health departments do not have the technical resources to build and maintain real-time electronic disease surveillance systems. Our application service provider model has been much more successful, and we now recommend that states form coalitions to share the costs of such services.

The project very early identified the need for a computing component to reside within the firewall of a health system, connected to the hospital's HL7 message router. This component would function as a case detector in a distributed public health surveillance scheme linking laboratory and radiology data to increase the specificity of case detection. It has proven very difficult to disseminate this technology, perhaps due to the complexity of the idea. Nevertheless, the threat of bioterrorism has created a need for such technology, and this approach, or something with equivalent function, must be deployed.

Adherence to NEDSS architectural standards was an early design objective that we have met. RODS 1.5 closely follows NEDSS architectural, software, messaging, and data specifications. Our success is a strong validation of those standards. We will gain further understanding of the standards as we attempt to use RODS components including HL7 listeners, natural language parsers, message parsers, databases, user interfaces, notification subsystems, and detection algorithms with other NEDSS compliant systems. An ongoing project will use RODS to collect chief complaints and integrate them into the Utah Department of Health's planned NEDSS system.

We have demonstrated the ability to rapidly deploy RODS in a special event with the added advantage that the system persisted after the event. This experience suggests strongly that RODS or similar systems be considered an alternative to drop-in surveillance.

Our future plans are to meet our initial design objective to develop early-warning capability for a large, outdoor release of anthrax, especially ensuring that the data and analysis produced by RODS are reviewed by public health. This goal will require improvements in the interfaces and the detection algorithms to reduce false alarms and to vastly improve the efficiency with which anomalies are evaluated by use of multiple types of data, better interfaces, and implementation of the look-back function. We would like to enlarge as quickly as possible the application service provider to include more states and more types of clinical data so that states will be in a position to prospectively evaluate the detection performance from different types of data on naturally occurring outbreaks.

Our long-term goals are to add additional disease scenarios to the design objectives such as detection of in-building anthrax release, vector-borne disease, food-borne disease, and a communicable disease such as severe acute respiratory syndrome (SARS).

Conclusion

RODS is a NEDSS-compliant public health surveillance system that focuses on real-time collection and analysis of data routinely collected for other purposes. RODS is deployed in two states and was installed quickly in seven weeks for the 2002 Olympics. Our experience demonstrates the feasibility of such a surveillance system and the challenges involved.

Outbreaks, emerging infections, and bioterrorism have become serious threats. It is our hope that the front-line of public health workers, astute citizens, and health care workers will detect outbreaks early enough so that systems such as RODS are not needed. However, timely outbreak detection is too important to be left to human detection alone. The notion that public health can operate optimally without timely electronic information is as unwise as having commercial airline pilots taking off without weather forecasts and radar.

References ■

1. Kaufmann A, Meltzer M, Schmid G. The economic impact of a bioterrorist attack: are prevention and postattack intervention programs justifiable? Emerg Infect Dis. 1997;3(2):83–94.
2. Guillemin J. Anthrax: the investigation of a deadly outbreak. N Engl J Med. 2000;343:1198.
3. Siegrist DW. The threat of biological attack: why concern now? Emerg Infect Dis. 1999;5:505–8.
4. Roush S, Birkhead G, Koo D, Cobb A, Fleming D. Mandatory reporting of diseases and conditions by health care professionals and laboratories. JAMA. 1999;282:164–70.
5. Ashford DA, Kaiser RM, Bales ME, et al. Planning against biological terrorism: lessons from outbreak investigations. Emerg Infect Dis. 2003;9:515–9.
6. Silver J. Local doctors fail their test on diagnosing germ terrorism. Pittsburgh Post-Gazette. 2000;February 13. Available at: http://www.post-gazette.com/healthscience/2000213biowar3.asp. Accessed July 13, 2003.
7. Aston C. Biological warfare canaries [biological attack detection]. IEEE Spectrum. 2001;38(10):35–40.
8. Ackelsberg J, Layton M. Update #5: Terrorist Attack at the World Trade Center in New York City: Medical and Public Health Issues [online] 2001. <http://www.nyc.gov/html/doh/html/cd/wtcf.html>. Accessed May 16, 2003.
9. Wagner MM, Aryel R, Dato V. Availability and Comparative Value of Data Elements Required for an Effective Bioterrorism Detection System. Washington, DC: Agency for Healthcare Research and Quality, 2001.
10. Lewis MD, Pavlin JA, Mansfield JL, et al. Disease outbreak detection system using syndromic data in the greater Washington DC area. Am J Prev Med. 2002;23:180–6.
11. Lober WB, Thomas Karras B, Wagner MM, et al. Roundtable on bioterrorism detection: information system-based surveillance. J Am Med Inform Assoc. 2002;9:105–15.
12. Tsui F-C, Espino JU, Wagner MM, et al. Data, network, and application: technical description of the Utah RODS Winter Olympic Biosurveillance System. Proc AMIA Symp. 2002:815–9.
13. Paulson T. Region alert to bioterror, but health-care system underfunded [online] 2001. <http://seattlepi.nwsource.com/local/40829_bio29.shtml>. Accessed March 6, 2002.
14. Pueschel M. DARPA System Tracked Inauguration For Attack [online] 2001. <http://www.usmedicine.com/article.cfm?articleI D=172&issueID=25>. Accessed March 6, 2002.
15. National Electronic Disease Surveillance System (NEDSS): a standards-based approach to connect public health and clinical medicine. J Public Health Manag Pract. 2001;7(6):43–50.
16. NEDSS systems architecture. April 15, 2001. Available at: http://www.cdc.gov/nedss/nedssarchitecture/nedsssysarch2.0.pdf. Accessed July 13, 2003.
17. Thacker S, Berkelman R. Public health surveillance in the United States. Epidemiol Rev. 1988;10:164–90.
18. DoD-GEIS. Electronic Surveillance System for Early Notification of Community-based Epidemics (ESSENCE) [online] 2003. <http://www.geis.ha.osd.mil/GEIS/SurveillanceActivities/ESSENCE/ESSENCE.asp>. Accessed May 16, 2003.
19. Schafer, K. LEADERS (Lightweight Epidemiology Advanced Detection & Emergency Response System) [online] 2001. <http://www.tricare.osd.mil/conferences/2001/agenda.cfm>. Accessed May 10, 2001.
20. Zelicoff A, Brillman J, Forslund D, et al. The Rapid Syndrome Validation Project (RSVP) [online] 2001. <http://www.cmc.sandia.gov/bio/rsvp/SAND%20No.pdf>. Accessed May 17, 2003.
21. Tsui F-C, Wagner MM, Dato V, Chang C-CH. Value of ICD-9-coded chief complaints for detection of epidemics. Proc AMIA Symp. 2001:711–5.
22. Ivanov O, Wagner MM, Chapman WW, Olszewski RT. Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. Proc AMIA Symp. 2002:345–9.
23. Wagner MM, Robinson JM, Tsui F-C, Espino JU, Hogan WR. Design of a national retail data monitor for public health surveillance. J Am Med Inform. 2003;10:409–18.
24. Gesteland PH, Gardner RM, Tsui F-C, et al. Automated syndromic surveillance for the 2002 Winter Olympics. J Am Med Inform. 2003;10:(in press).
25. Liu Z, Tsui F-C, Zeng X. Cache table design for disease surveillance system. Proc AMIA Symp. 2002:1086.
26. Java 2 Platform Enterprise Edition (J2EE) [online] 2003. <http://java.sun.com/j2ee/>. Accessed February 10, 2003.
27. Olszewski RT. Bayesian classification of triage diagnoses for the early detection of epidemics. Proc 16th Int FLAIRS Conference. 2003:412–6.
28. Orfanidis SJ. Optimum Signal Processing (ed 2). New York: McGraw-Hill, 1988.
29. Wong W, Moore A, Cooper G, Wagner M. Rule-based anomaly pattern detection for detecting disease outbreaks. Proceedings of the Conference of the American Association of Artificial Intelligence (AAAI); 2002.
30. Panackal AA, M'ikanatha NM, Tsui F-C, et al. Automatic electronic laboratory-based reporting of notifiable infectious diseases. Emerg Infect Dis. 2001;8:685–91.