Chapter 8

RODS

The Real-time Outbreak and Disease Surveillance (RODS) system was initiated by the RODS Laboratory at the University of Pittsburgh in 1999. The system is now an open source project under the GNU license. The RODS development effort has been organized into seven functional areas: overall design, data collection, syndrome classification, database and data warehousing, outbreak detection algorithms, data access, and user interfaces. Each functional area has a coordinator for the open source project, and there is an overall coordinator responsible for the architecture, overall integration of components, and overall quality of the JAVA source code. Figure 8-1 illustrates the RODS' system architecture.

The RODS system as a syndromic surveillance application was originally deployed in Pennsylvania, Utah, and Ohio. As of 2006, RODS performs emergency department surveillance for other states of California, Illinois, Kentucky, Michigan, New Jersey, Nevada, and Wyoming through an ASP model at the University of Pittsburgh, and through local installations in Taiwan, Canada, Mississippi, Michigan, California, and Texas. As of June 2006, about 20 regions with more than 200 healthcare facilities connected to RODS in real-time. It was also deployed during the 2002 Winter Olympics (Espino et al., 2004). It also serves as the user interface for national over-the-counter medication sales surveillance data collected through the NRDM.

The conceptual architecture of the RODS system is shown in Figure 8-1. Multiple data sources are collected and stored in a database and data warehouse where they are made available to outbreak algorithms and the RODS user interface.
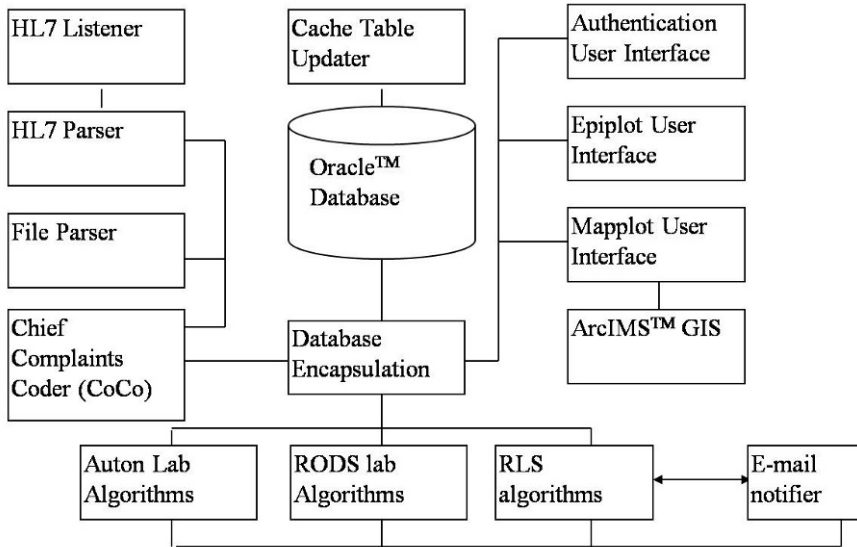
Figure 8-1. RODS system architecture (Espino et al., 2004).

The latest version of RODS system is RODS 6. RODS 6 is built from the ground up to be pluggable and part of a larger biosurveillance system as a biosurveillance grid node that incorporates as well as offers services. New data types and algorithms can be easily incorporated into the system without the need for database redesign or coding of the core software. RODS 6 also provides a robust API so that external applications can leverage the data collection, visualization, and data analysis capabilities of RODS.

1.   RODS DATA COLLECTION

RODS collects healthcare registration data in real time from participating hospitals via a standard called HL7. Specifically, healthcare registration data consist of the age, gender, home zipcode, date/time of admission, and a free-text chief complaint of the patient.

The National Retail Data Monitor (NRDM) is a component of the RODS system, collecting and analyzing daily sales data for OTC medication sales. It also collects and analyzes chief complaints data from various hospitals. NRDM monitors more than 29,000 retail stores including stores from 12 big chains in the US and its territories for OTC medication sales 24 hours a day/ 7 days a week as of May 2009 (a screenshot of its deployment around the US is shown in Figure 8-2). Daily batch feeds of sales data from those stores are

received by NRDM everyday by midnight. Individual medication sales data are aggregated into one or more of the 18 OTC categories (Table 8-1) before being aggregated spatially by zip code according to store location.

There are plans to integrate laboratory orders, dictated radiology reports, dictated hospital reports, and poison control center calls in future versions.

The RODS system currently monitors 7 healthcare registration prodrome categories, as shown in Table 8-2.

Table 8-1. Eighteen over-the-counter medication categories monitored by NRDM.

|	
Antidiarrheal	Cough syrup adult liquid
Antifever pediatric	Cough syrup adult tablet
Antifever adult	Cough syrup pediatric liquid
Bronchial remedies	Electrolytes pediatric
Chest rubs	Hydrocortisones
Cold relief adult liquid	Nasal product internal
Cold relief adult tablet	Thermometers
Cold relief pediatric liquid	Throat lozenges
Cold relief pediatric tablet	Others
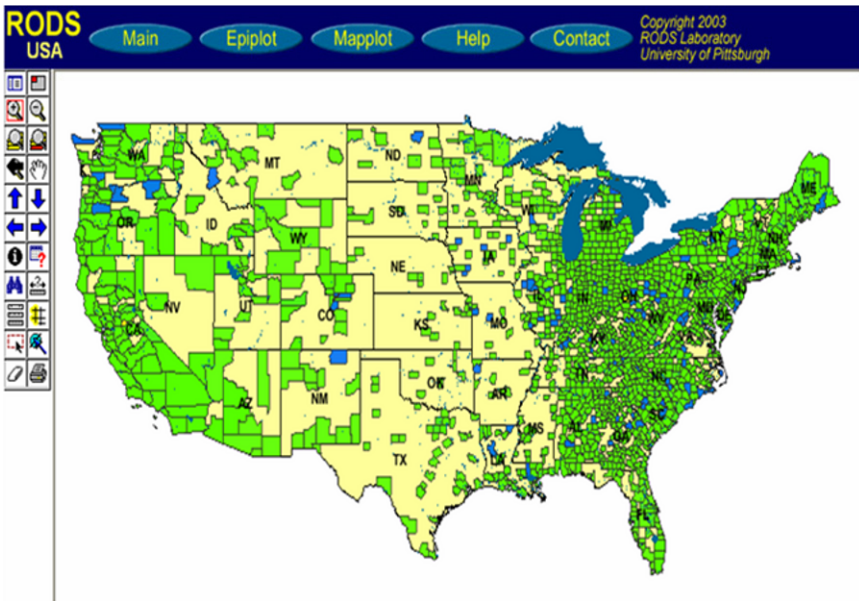
Figure 8-2. NRDM deployment at 20,000 stores as of 2002 (Wagner et al., 2003).

Table 8-2. Syndrome categories monitored by RODS.

Gastrointestinal	Rash
Hemorrhagic illness	Respiratory
Constitutional	Botulism-like/botulism
Neurologic	

The RODS data are collected in real-time through HL7 messages from other computer systems such as registration systems and laboratory information systems, over a Secure Shell–protected Internet connection in an automated mode.

2.   RODS DATA ANALYSIS

One of the major strengths of RODS is in data analysis. Several syndrome classification approaches have been tested and implemented in the RODS system. It applies a keyword classifier and an ICD-9 classifier to chief complaint data. The CoCo module, a syndrome mapping component, has been tested in multiple settings (Olszewski, 2003). For the respiratory syndrome, based on manually-classified results, CoCo achieves a 77% sensitivity level and 90% specificity level (Wagner et al., 2004b). The classifier's performance for other syndrome categories can also be found in (Wagner et al., 2004b). Chapman et al. (2005) proposed a Bayesian network-based semantic model, which has shown to classify free-text chief complaints effectively at the expense of added system complexity and computational overhead. The performance of the classifier represented by the ROC curve for each syndrome category varies between 0.95 and 0.99.

The RODS laboratory, in collaboration with the Auton Lab at Carnegie Mellon University, continues to develop additional algorithms to model both the temporal fluctuations and spatial distribution patterns in syndromic surveillance datasets. The current open source release of the RODS system includes implementations of several on-the-fly outbreak detection algorithms: wavelet-detection algorithms, Moving Average, CUSUM with Exponentially Weighted Moving Average, and Recursive Least Square (RLS). Methods including SMART, scan statistics, and WSARE are also being developed and tested. A future release will allow the import and export of data as common text files such that stand-alone algorithms and statistical software packages can be used to analyze the data.

CUSUM and SMART are also used in the BioSense system. They were discussed in the previous section. What's Strange About Recent Events (WSARE) algorithm (Wong et al., 2003, 2005) evaluates all the possible rules that are made up of any data feature components (e.g., a two-component

rule could be Gender = Male and Home = NW) in both recent data and baseline data. The rules that have the largest discrepancy of the proportions between the recent data and baseline data are detected as rules summarizing the most significant patterns of anomalies. WSARE 3.0 has been evaluated retrospectively using the data from the Israel Center for Disease Control and has shown its capability of detecting the outbreak on the second day from its onset (Kaufman et al., 2005).

In the latest release of RODS 6, high-fidelity injection detectability experiments (HiFIDE) are integrated for outbreak simulation and algorithm testing. HiFIDE enable public health officials to analyze the detectability characteristics of a surveillance system operating in their jurisdiction. HiFIDE inject synthetic outbreak data (spikes) into real surveillance data from a particular jurisdiction. The HiFIDE spike is both high-fidelity in contour and in scale by first deriving the spike shape and then rescaling the spike from real surveillance data collected during outbreaks that occurred in other regions. In particular, its interface focuses on depicting the expected sensitivity, specificity, and timeliness of detection for outbreaks of varying sizes, etiologies, and geographic and demographic scopes. A HiFIDE window is shown in Figure 8-3.
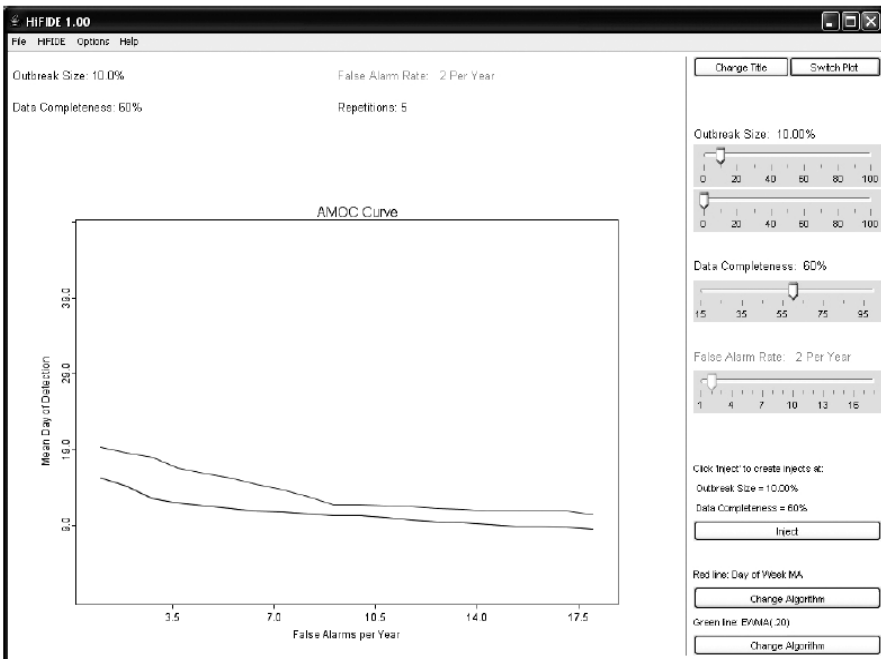

Figure 8-3. Sensitivity plots in HiFIDE (Wallstrom et al., 2005).

3.   RODS VISUALIZATION, INFORMATION DISSEMINATION, AND REPORTING

The RODS system provides multiple graphing techniques with both time-series and geographical displays available via an encrypted, password-protected Web interface. Three different data views – Main, Epiplot, and Mapplot – are supported. Figures 8-4 to 8-6 are three example views of RODS user interface. These views are implemented using JFreeChart (an open-source graphing package) and ArcIMS (an Internet GIS server developed by the Environmental Systems Research Institute, Inc.).

The RODS Main screen (Figure 8-4) shows time-series plots updated on a daily basis for each syndrome. The intention of the Main screen is that of a "threat" board in a situation room. The Main screen refreshes itself every two minutes if left displayed. The graphs can be plotted with different event monitoring algorithms such as moving average and CUSUM. The user can choose to view these plots by county or for the whole state.

The RODS Main screen is limited to viewing six OTC (individual types of medication and prodrome categories cannot be selected from the OTC Main screen) or healthcare registration charts for the last seven days, whereas Epiplot screen (Figure 8-5) is highly interactive. EpiPlot allows the user to specify the syndrome, particular geographic region, start dates, and end dates, to
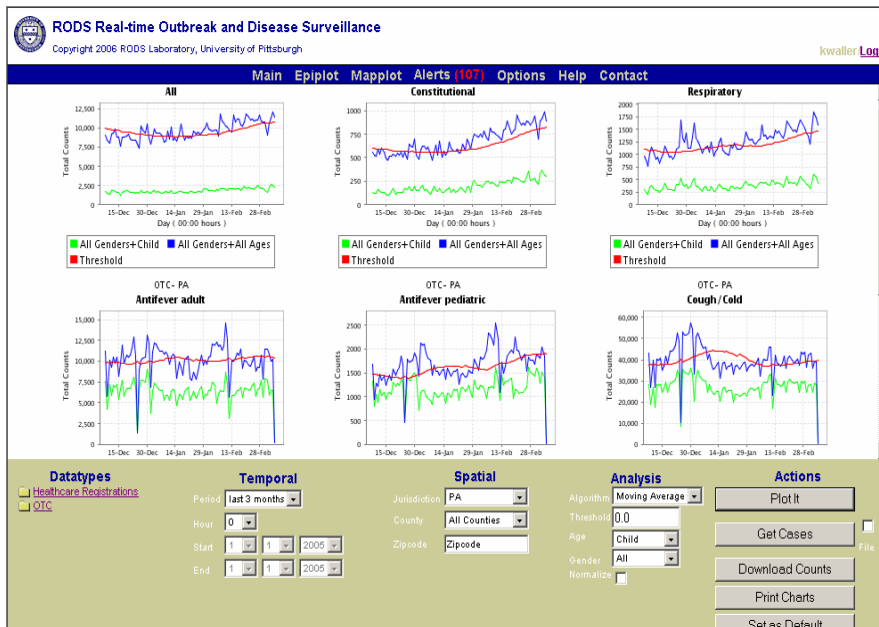

Figure 8-4. RODS system Main screen (source: RODS Laboratory, University of Pittsburg).

generate customized time-series plots. Users can choose to analyze the data using one of four on-the-fly analysis algorithms – CuSUM with EWMA, RLS, Wavelet or Moving Average. A "get cases" button allows users to view case-level detail for encounters making up the specific time-series.
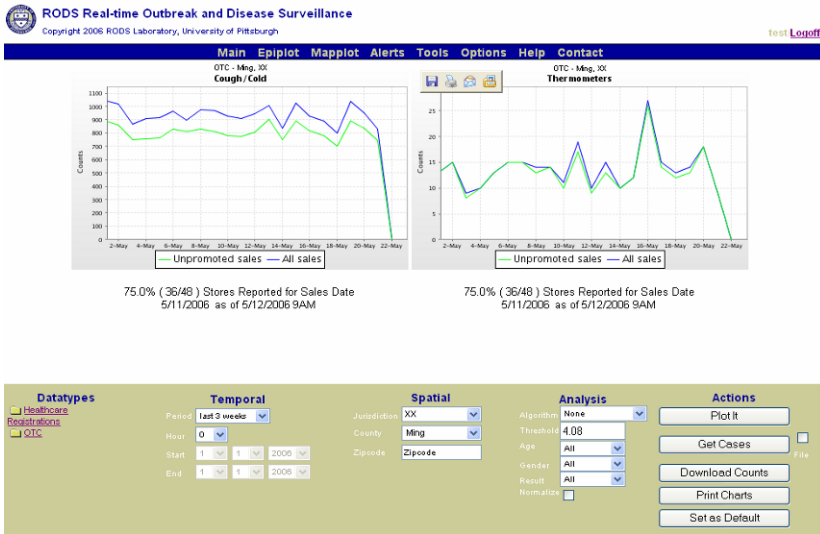

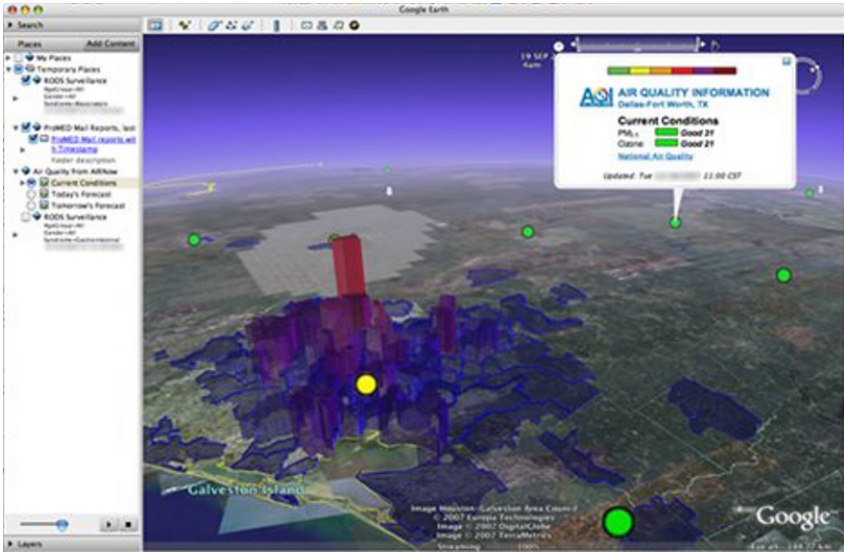*Figure 8-5*. RODS Epiplot screen (source: RODS User Manual).

Figure 8-6. Mapplot output displayed in Google Earth (source: the RODS project at Source Forge.net).
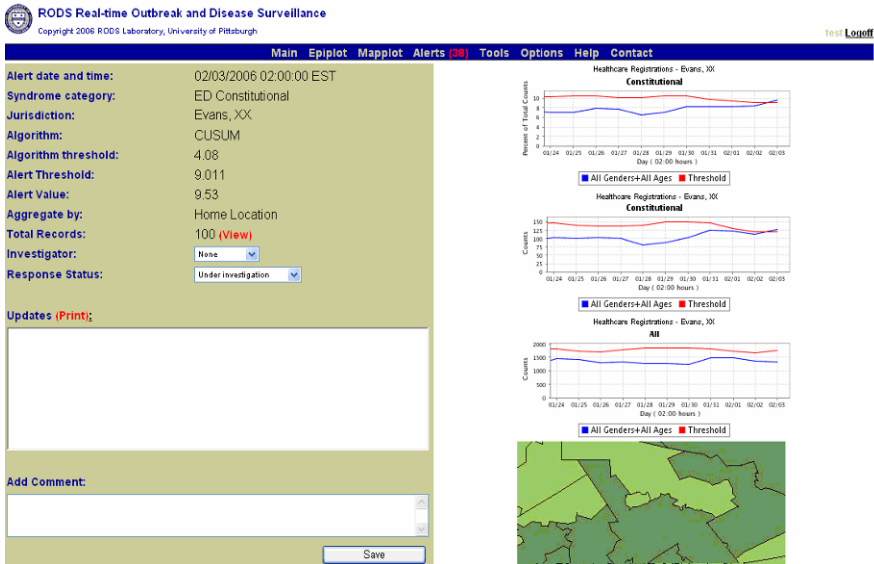
Figure 8-7. RODS alerts (monitoring healthcare registrations only) (source: RODS manual).

The Mapplot screen provides an interface to the ArcIMS package, to display disease cases' spatial distribution using patients' zip code information. Figure 8-6 is a Mapplot screenshot of Google Earth geographic view of daily frequencies of one type of OTC sales.

The Alerts page (Figure 8-7) provides detailed information about each alert for a defined jurisdiction. An alert is registered each time data analyzed exceeds the thresholds set by one or more of the four algorithms in use.

4. CASE STUDY: SYNDROMIC SURVEILLANCE WITH RODS FOR THE 2002 WINTER OLYMPICS

RODS was deployed at the 2002 Winter Olympics in Salt Lake City for bioterrorism and public health surveillance. The main purpose of implementing RODS was to automate an otherwise expensive, round-the-clock surveillance process. It was a successful test of RODS deployment in such short-term drop-in situations (Gesteland et al., 2003).

During the Olympics, encounter data were collected from 19 urgent care centers and nine emergency departments owned and operated by Intermountain Health Care (IHC), University of Utah Health Sciences Center (UUHSC) and from the University of Utah Hospital's emergency department and the

Polyclinic located in the Olympic Village. Together these emergency rooms and urgent care centers serve about 70% of the population of Utah.

RODS takes advantage of existing HL7 message routers in healthcare systems to receive admission, discharge, and transfer (ADT) data in real time from clinical information systems. HL7 message routers consist of HL7 data listeners and HL7 parsers. The HL7 listeners establish TCP/IP connections between RODS and IHC and UUHSC. The HL7 parser uses regular expressions to parse each data segment in an HL7 message. The parsed ADT messages are centralized into an Oracle8i database for data retrieval and analysis. Figure 8-8 shows a sample HL7 message from one of health systems. The primary keys for a HL7 ADT message contain sending facility, ADT message type, medical record number, patient class, and visit number (Tsui et al., 2003).


```
MSH|^~\&|HELP|xxx|COMMON|EXTERNAL|200202241715||ADT
^A04|2002022XXXXXXXX|P|2.3<CR>
PID|||123456789||||^020|M|||^^^^84204|||||<CR>
PV1||E|||||||||||||||||98765432|||||||||||||||||||
||||200202XXXXXXX|||<CR>
DG1||||SORE THROAT,COUGH<CR>
IN1||||||||||||||||||||||||||||||||||||||||||||||^^^^
84056<CR>
<ETX>
```

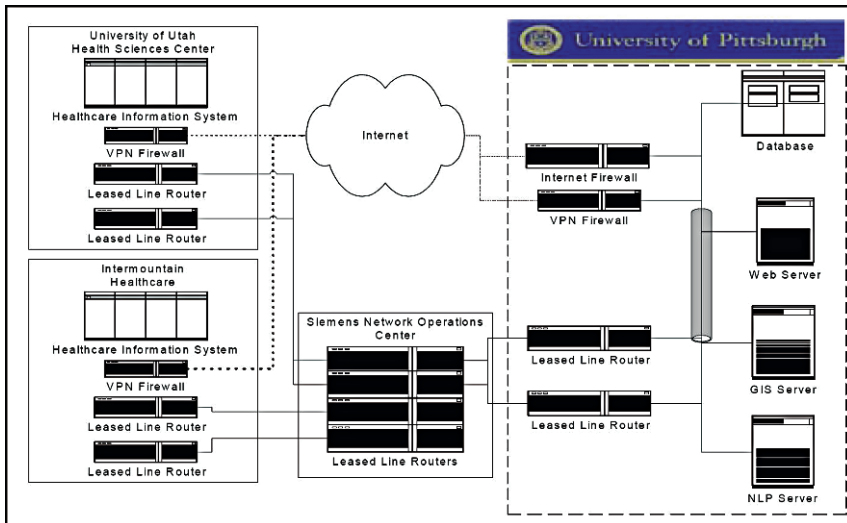*Figure 8-8.* Sample HL7 ADT messages (Tsui et al. 2003).

Figure 8-9. Network architecture of RODS implementation in Utah (Tsui et al. 2003).

The HL7 messages were communicated through a secure network infra-structure between Utah RODS and the data providers consisting of virtual private networks (VPN) and leased lines (Figure 8-9). Siemens Medical Systems (SMS) contributed to the initial establishment for the site-to-site VPNs. Utah RODS processes run on dedicated databases and servers including a Web server, a GIS server, and a natural language processing server, protected by Internet firewalls residing on separate servers.

To classify patients into a prodrome category utilizing free-text chief complaints as input, Utah RODS uses two Natural Language Processors (NLPS) – Bigram 8 and PLUSI0. They map a chief complaint into one of seven prodromes-respiratory, diarrheal, botulinic, viral, encephalitic, hemorrhagic, and rash. Bigram is a simple NLPS, developed at the University of Pittsburgh, computing the probability of a specific prodrome category of a patient based on pair of words in a free-text chief complaint. PLUS was developed at the University of Utah. PLUS classifies a free-text chief complaint using a more sophisticated Bayesian network. Both NLPS operate in real time using client-server TCP/IP socket connections. Whenever a chief complaint is available for processing, the RODS server sends a message to PLUS on the NLP server, and it returns the classification of the case based on the free-text chief complaint (Tsui et al., 2003).

RODS analyzes the data for anomalous densities of cases compared with historical patterns. The analyses were conducted every 4 h, and were frequently visually inspected by RODS staff through RODS user interfaces. The primary statistical tool used by RODS during the Olympics for automated pattern recognition was the RLS adaptive filter. RLS (as discussed in Chapter 4) computes an expected count of each syndrome within a region from historical data, adjusting its model coefficients based on prediction errors. RLS algorithm has an advantage over other potential algorithms in such "drop-in" situations because it requires only a few days of historical data to generate model co-efficients. The WSARE algorithm was also used to perform heuristic searches over combinations of temporal and spatial features to detect anomalous densities of cases in space and time.

Over 114,000 acute care encounters were monitored between February 8 and March 31, 2002. The RODS system signaled two alarms; both times the appropriate authorities were notified and the alarms were determined to be false positives.

At the Olympics, the largest problems faced by the investigators corres-ponded to data sharing. The major data contributors (IHC and University of Utah) could not share the same HL7 data sets because of proprietary data collection issues. It slowed down the process of implementation in a situation where time was essential. RODS project spent a considerable amount of time during the 7 weeks in this project managing administrative issues instead of actually setting up the RODS system. Despite the inherent limitations of the 7-week establishment of the RODS system, the project was highly successful in

proving how a computer-based, minimally invasive syndromic surveillance system can work.

5.   FURTHER READINGS

We provide the following project link and some key readings for the readers who might be interested in learning more details about the RODS system.

Project link:

https://www.rods.pitt.edu/site/

Important readings:

1. Wu, T. S., F. Y. Shih, M. Y. Yen, J. S. Wu, S. W. Lu, K. C. Chang, C. Hsiung, J. H. Chou, Y. T. Chu, H. Chang, C. H. Chiu, F. C. Tsui, M. M. Wagner, I. J. Su, and C. C. King (2008), "Establishing a nationwide emergency department-based syndromic surveillance system for better public health responses in Taiwan," *BMC Public Health*, 8, p 18.
2. Shen, Y., C. Adamou, J. N. Dowling, and G. F. Cooper (2008), "Estimating the joint disease outbreak-detection time when an automated biosurveillance system is augmenting traditional clinical case finding," *Journal of Biomedical Informatics*, 41, pp 224–231.
3. Wallstrom, G. L., and W. R. Hogan (2007), "Unsupervised clustering of over-the-counter healthcare products into product categories," *Journal of Biomedical Informatics*, 40(6), pp 642–648.
4. Dara, J., J. N. Dowling, D. Travers, G. F. Cooper, and W. W. Chapman (2007), "Evaluation of preprocessing techniques for chief complaint classification," *Journal of Biomedical Informatics*, 41(4), pp 613–623.
5. Espino, J. U., M. M. Wagner, F. C. Tsui, H. D. Su, R. T. Olszewski, Z. Lie, W. Chapman, X. Zeng, L. Ma, Z. W. Lu, and J. Dara (2004), "The RODS Open Source Project: removing a barrier to syndromic surveillance," *Medinfo*, 11(Pt 2), pp 1192–1196.
6. Tsui, F.-C., J. U. Espino, M. M. Wagner, P. Gesteland, O. Ivanov, R. T. Olszewski, Z. Liu, X. Zeng, W. Chapman, W. K. Wong, and A. Moore (2002), "Data, network, and application: technical description of the Utah RODS Winter Olympic Biosurveillance System." *Proceedings of the AMIA Symposium*, pp 815–819.