Chapter 7

BIOSENSE

BioSense is part of the US CDC's Public Health Information Network (PHIN) framework managed through the CDC BioIntelligence Center. It supports early outbreak detection at the local, state, and national levels, by monitoring the size, location, and rate of spread of an outbreak; monitoring seasonal trends of influenza and other disease indicators; and assisting in case-finding for epidemiologic investigations.

In March 2005, BioSense had more than 340 state and local health department user accounts, representing 49 states. Its user base continues to expand. The current implementation status of BioSense (as of June 2008) is shown in Figure 7-1. The system has also been used in several high-profile events (e.g., the G8 meeting in 2004) (Bradley et al., 2005; Ma et al., 2005; Sokolow et al., 2005).

Figure 7-2 shows the BioSense system architecture. Specifically, BioSense consists of the following system components (BioSense, 2008):

- Data Transmission: assuring the secure, timely, and routine receipt of health data for public health surveillance. BioSense requires data to be transmitted over the PHIN Messaging System (PHINMS). PHINMS is an interoperable messaging system developed by CDC for data providers to transmit private data either as standardized messages and vocabulary securely over the Internet in real-time or in batches.
- Data Analysis: establishing a set of statistical methods and tools to assist public health analysts to detect potential public health events and make informed decisions. At the CDC BioIntelligence Center (BIC) each day, the public health analysts monitor, analyze, and interpret facility, state, and national trends or anomalies in the BioSense data and provide further analytic and reporting support to state and local public health departments.

H. Chen et al., *Infectious Disease Informatics: Syndromic Surveillance for Public Health and BioDefense*, Integrated Series in Information Systems 21, DOI 10.1007/978-1-4419-1278-7_7,

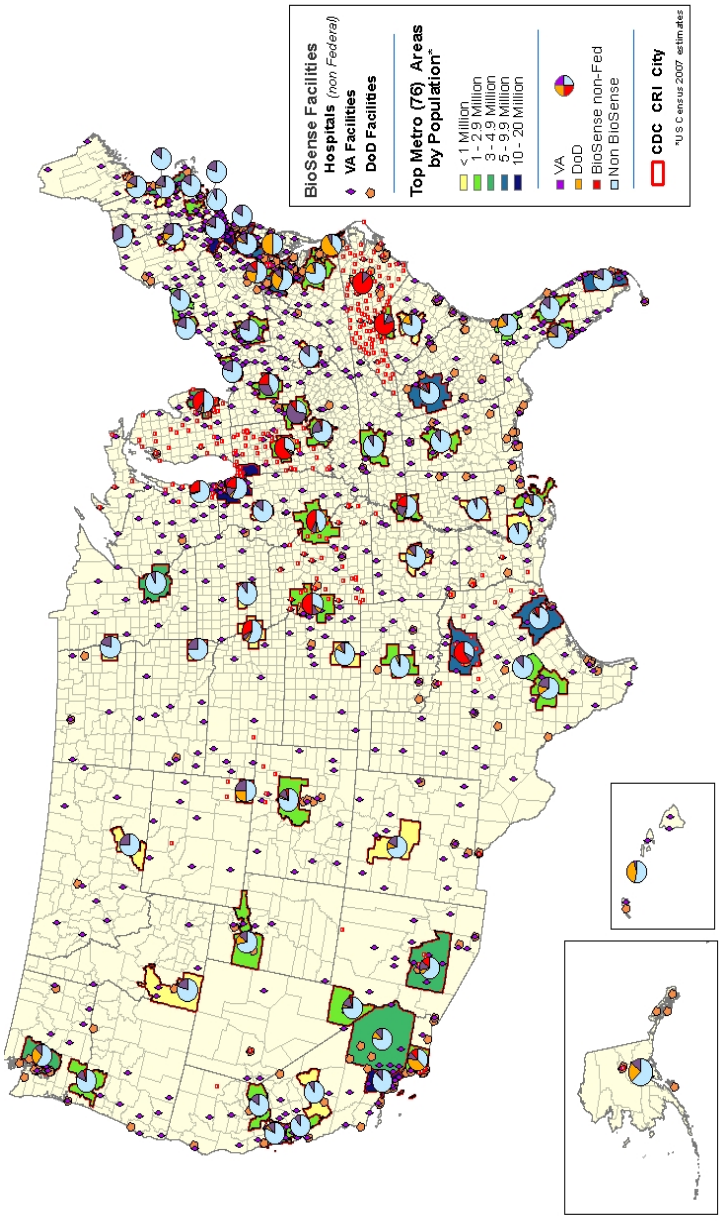


Figure 7-1. BioSense participation in top 76 MSAs as of June 2008 (BioSense, 2008).

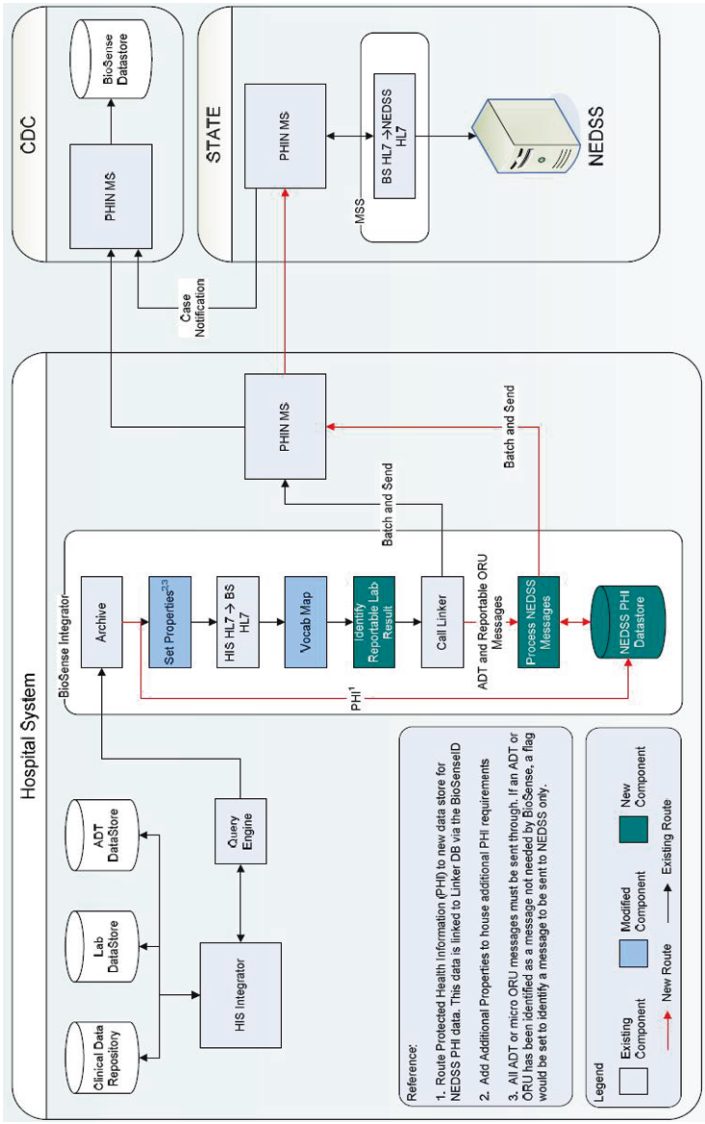


Figure 7-2. BioSense with ELR reporting intergration (source: BioSense Web site).

- Data Reporting: on a near real-time basis, providing useful views of data through time-series graphs and geospatial maps, for state and local public health as well as for CDC program staff.
- Public Health Response: providing state and local public health staff real-time access to existing data from healthcare organizations, state syndromic surveillance systems, national laboratories, and other data sources for investigations, outbreak responses, and public health interventions.

Figure 7-2 also shows the recent collaborative efforts between CDC's National Electronic Disease Surveillance System (NEDSS) and BioSense. The goal of the collaboration is to establish interoperable communications between a hospital system to a state-based electronic disease surveillance system (e.g., NEDSS Base System or any NEDSS compliant system) consistent with CDC PHIN standards.

1. BIOSENSE DATA COLLECTION AND PREPROCESSING

BioSense data providers include Department of Defense (DoD)-Military Treatment Facilities (MTF), the Department of Veterans Affairs (VA), the Laboratory Response Network (LRN), and Electronic Laboratory Results (ELR) reporting systems. The system accepts, receives, and collects up to four ICD-9-CM diagnosis codes identifying the reasons for ER visits and procedure-encoded CPT ordered for every ambulatory care visit from DoD-MTF and VA. Clinical laboratory test orders are collected nationally through the commercial lab operator LabCorp (Laboratory Corporation of America). It also receives lab results from BioWatch environmental sensors (Sokolow et al., 2005). BioSense supports automated messaging through HL7 protocols in either a batch mode or a near real-time mode. The data types BioSense collects from hospital EDs and ambulatory care include patient chief complaint, physician diagnosis, supporting patient demographic data, daily hospital census, ED-specific clinical data, microbiology test orders and results, radiology orders and results, and medication orders.

The 11 syndrome categories monitored by BioSense are shown in Table 7-1. To allow surveillance of more granular events than is possible using the 11 syndromes, BioSense medical expert staff developed 78 more subsyndromes. These subsyndrome definitions can be found at the BioSense project Web site (CDC, 2007).

Table 7-1. Eleven syndrome categories monitored by BioSense.

Fever	Neurologic
Gastrointestinal	Rash
Hemorrhagic illness	Severe illness and death
Localized cutaneous lesion	Specific infection
Lymphadenitis	Respiratory
Botulism-like/botulism	

Data in ICD-9-CM form are mapped to 11 syndromes based on a mapping schema created in 2003 by a multiagency working group (CDC, 2007). Free-text data are mapped to subsyndromes using the text word search. Most keywords in the chief complaint to subsyndrome mapping table were derived from the EARS system Text String Search method. It contains both English and certain Spanish keywords and includes regular terms, misspellings, word fragments, and abbreviations. The mapping is continually improving the keyword search list by examining the original free text and its corresponding mapping results. Keywords were modified during the initial implementation period. The majority of the keywords in the free-text physician diagnosis to subsyndrome mapping table were derived from terms that appeared in ICD-9-CM descriptions (CDC, 2007). At the same time, BioSense employs a Bayesian classifier – CoCo from the RODS laboratory – for syndrome classification.

2. BIOSENSE DATA ANALYSIS

BioSense uses the CUSUM algorithm for anomaly detection. The CUSUM algorithm is used as a short-term surveillance technique to indicate recent data changes through the comparison of moving averages (Bradley et al., 2005). Because of the high variability within the data, CUSUM values are computed for each date-source-syndrome combination at the state or metropolitan reporting area (MRA) level rather than for individual ZIP codes (Bradley et al., 2005).

The other detection algorithms available from BioSense include EWMA and SMART. EWMA and SMART algorithms are also used to predict the day-source-syndrome counts at the ZIP code level, with seasonality and day-of-week effects considered. The calculations are conducted on a daily basis. Spatial-temporal clustering methods such as various scan statistics are also being explored by the BioSense system. BioSense explored the use of SaTScan with a separate run for each month to detect spatial disease clusters. SaTScan is set to scan a maximum circle radius of 100 km with each ED facility as one geographic unit. Poisson probability model is used to model the disease

rates, and clusters are identified by locating the geographic areas that do not conform to these model-predicted disease rates.

3. BIOSENSE DATA VISUALIZATION, INFORMATION DISSEMINATION, AND REPORTING

BioSense is an Internet-accessible, secure system. It displays data in multiple formats including line graphs, maps, tabular summaries, and case details. Graph plotting for individual data source, individual syndrome category,

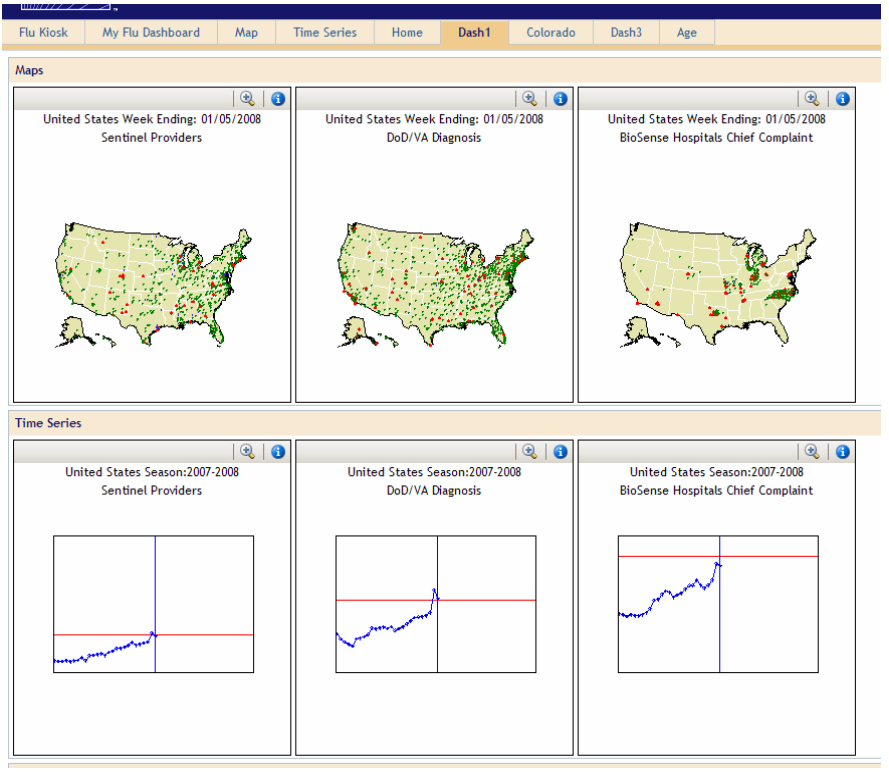


Figure 7-3. BioSense Influenza tool that merges multiple sources (source: BioSense Web site).

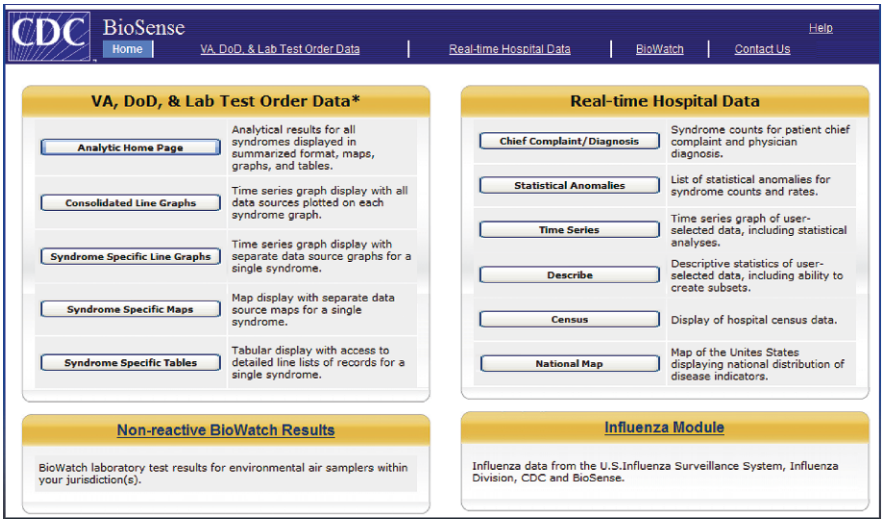
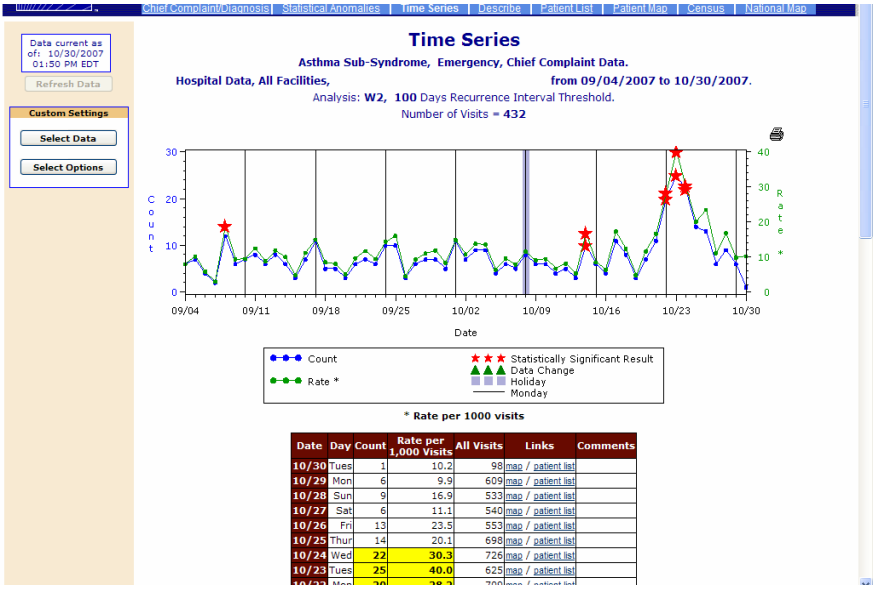


Figure 7-4. BioSense homepage showing available surveillance functionalities (source: BioSense Web site).

and different level of geographical regions is also available (Figure 7-3). On its homepage, as shown in Figure 7-4, it provides a collection of analysis and visualization functionalities. For VA, DoD, and Lab Test Order Data, (1) it can display time series graphs or map graphs of all data sources for each syndrome or a selected specific syndrome (the example of asthma time-series is shown in Figure 7-5); (2) it has tabular display with access to detailed line lists of records for a single syndrome; (3) infection alerts for several bioterrorism agents can also be reported. For real-time hospital data, a line list of statistical anomalies found by BioSense analysis, time series and map display for syndrome counts, and as well as drill-down patient details are all available.

CDC BioIntelligence Center is the agency responsible for monitoring anomalies detected by BioSense. The lightweight directory access protocol (LDAP) is employed for information reporting.



Date	Day	Count	Rate per 1,000 Visits	All Visits	Links	Comments
10/30	Tues	1	10.2	98	map / patient list	
10/29	Mon	6	9.9	609	map / patient list	
10/28	Sun	9	16.9	533	map / patient list	
10/27	Sat	6	11.1	540	map / patient list	
10/26	Fri	13	23.5	553	map / patient list	
10/25	Thur	14	20.1	698	map / patient list	
10/24	Wed	22	30.3	726	map / patient list	
10/23	Tues	25	40.0	625	map / patient list	

Figure 7-5. BioSense analysis page for Asthma query (source: BioSense Web site).

4. CASE STUDY: MONITORING HEALTH EFFECTS OF WILDFIRES USING BIOSENSE

From October 21 to October 26, 2007, wildfires spread across hundreds of thousands of acres of San Diego County, forcing the evacuation of more than 300,000 residents. During October 22–30, 2007, CDC personnel monitored BioSense for evidence of health effects possibly related to the wildfires in San Diego County.

In October 2007, data were being received from EDs at six of the 19 hospitals in San Diego County. These six hospitals were located near but outside the fire and evacuation areas (illustrated in Figure 7-6).

Data received by BioSense included age, sex, free-text patient-reported chief complaints, and diagnosis codes (usually ICD-9-CM codes). The first part of the standard procedure is syndrome classification. Diagnoses are assigned to one or more of the 11 general syndromes (shown in Table 7-1) and 78 more specific subsyndromes (e.g., asthma and dyspnea).

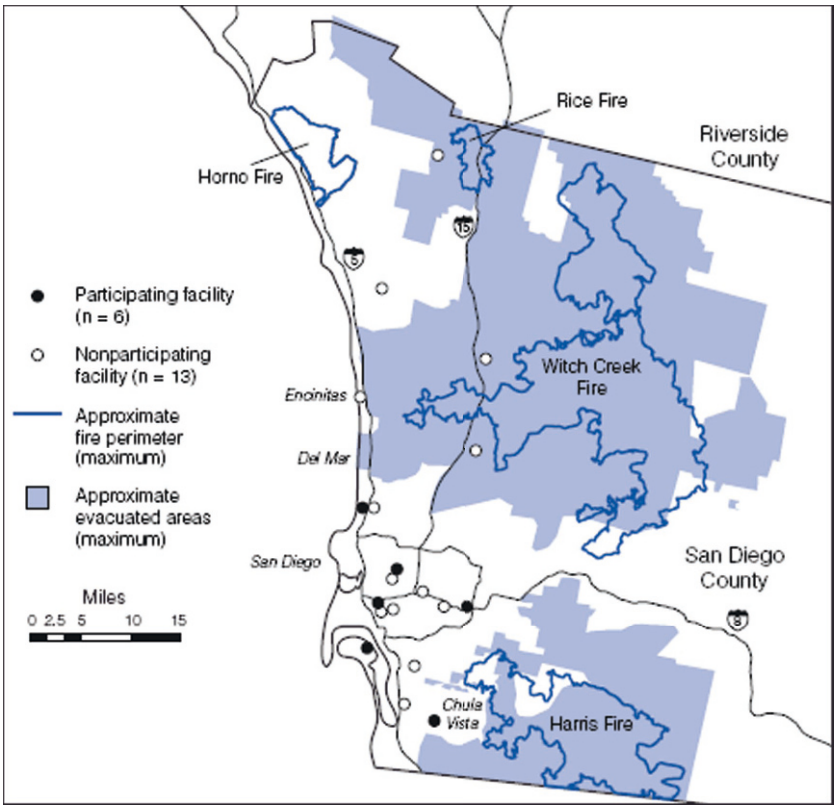
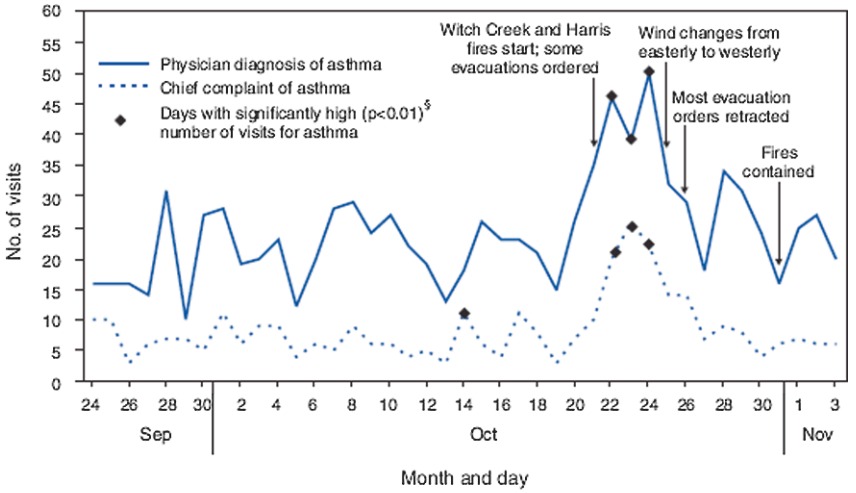


Figure 7-6. Hospital participation in BioSense, San Diego County, California, October 20-29, 2007 (Ginsberg et al., 2008).

These data are first centralized at CDC from hospital EDs. Within 2–3 hours, these data are processed at CDC and then made available in BioSense. The median time for chief complaints from patient visits to receipt of ED data at CDC is 8 hours. For diagnosis codes, the median time is 5 days.

For data analysis, the daily count of visits indicating diseases after the manual or automatic syndrome mapping is displayed on time-series graphs (Figure 7-7 shows an example time-series for counts of diagnoses and chief complaints of asthma) and compared with the predicted number based on a 7-day moving average. A modification of the EARS C-2 algorithm (Hutwagner et al., 2003) is used to determine statistical significance. A single-day visit count with a recurrence interval of $\geq$100 days (analogous to $p \leq 0.01$) is considered statistically significant.



* Free-text chief complaints are parsed for specified keywords and assigned to syndromes and subsyndromes.
† Based on *International Classification of Diseases, Ninth Revision, Clinical Modification* code 493 (asthma).
§ Statistical significance determined using a modification of the Early Aberration Reporting System (EARS) C-2 algorithm.

Figure 7-7. Time-series of ED visits by chief complaints and diagnosis of asthma – six participating hospitals, San Diego, California, September 22 – November 17, 2007 (Gingsberg et al. 2008).

During the wildfires, the BioSense system noted increases in total hospital visit volume and large increases in respiratory visits to hospitals, especially visits for asthma and dyspnea (difficulty in breathing/shortness of breath). The BioSense system detected significant ($p < 0.01$) increases in visits for asthma from October 22 to 24. When the winds shifted on October 25, asthma complaints and diagnoses began to decline.

BIC and San Diego County public health officials also worked together to conduct retrospective analyses of BioSense post-wildfire data. These analyses helped to gain a better understanding of how cardiovascular and respiratory diseases develop before, during, and after the fire and how patients with chronic respiratory illness were affected by exposure to the wildfire smoke. The collaboration between BIC and San Diego County public health officials proved to be useful and has led to increasing collaborative activities across CDC and with state and local public health officials. Lessons learned from this experience will help not only the next time wildfires strike, but also in other large-scale exposures to environmental hazards.

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 BioSense project.

Project link:

http://www.cdc.gov/BioSense/

Important readings:

1. BioSense working group. (June 2008) "BioSense Technical Overview of Data Collection, Analysis, and Reporting." Available at http://www.cdc.gov/BioSense/files/BioSense_Techn_Overview_102908_webpage.pdf
2. Ginsberg, M., J. Johnson, J. Tokars, C. Martin, R. English, G. Rainisch, W. Lei, P. Hicks, J. Burkholder, M. Miller, K. Crosby, K. Akaka, A. Stock, and D. Sugerman. (2008). "Monitoring Health Effects of Wildfires Using the BioSense System – San Diego County, California, October 2007." MMWR July 11, 2008.
3. Bradley, C. A., and H. Rolka, et al. (2005). "BioSense: Implementation of a National Early Event Detection and Situational Awareness System." MMWR (CDC) 54(Suppl), pp 11–20.
4. Sokolow, Leslie Z., N. Grady, H. Rolka, D. Walker, P. McMurray, R. English-Bullard, J. Loonsk. "Practice and Experience: Deciphering Data Anomalies in BioSense." MMWR August 26, 2005.
5. Ma, H., J. Tokars, R. English, T. Smith, C. Bradley, L. Sokolow, and H. Rolka. 2006 Jul 7. "Surveillance of West Nile Virus Activity Using Biosense Laboratory Test Order Data." Advances in Disease Surveillance [Online] 1:1.
6. R. English, P. McMurray, L. Sokolow, H. Rolka, D. Walker, J. Quinn III, and K. Cox. 2006 Jul 7. "Geographic Categorization Methods Used in BioSense." Advances in Disease Surveillance [Online] 1:1.