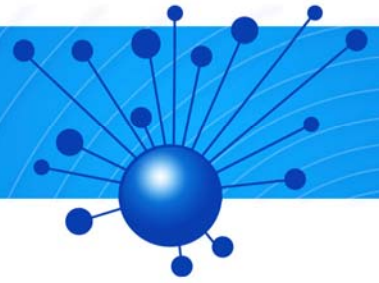


BioSense



***BioSense Technical Overview
of Data Collection, Analysis, and Reporting***

June 2008

Contents

Introduction.....	3
BioSense Background.....	3
BioSense Data Sources.....	3
BioSense Data Types.....	8
Access to BioSense.....	9
How BioSense Functions.....	9
PHIN Architecture.....	9
BioSense Operating Environment.....	10
Data Processing.....	11
Data Quality.....	12
Data Analysis Steps.....	12
Characterization of Data Anomalies.....	15
Data Pre-Processing.....	16
Making Decisions.....	17
How is the data used by the CDC.....	17
How is the data used by state and local public health and hospital facilities.....	17
Application Outputs.....	17
For More Information.....	22
Glossary.....	23

Figures

Figure 1. BioSense Participation in Top 76 MSAs as of June 2008.....	7
Figure 2. BioSense Data Connections.....	7
Figure 3. . BioSense Hospital Level A, B, C Direct Connections.....	8
Figure 4. BioSense homepage screen image.....	9
Figure 5. BioSense Technical Process Flow Diagram.....	10
Figure 6. BioSense Operating Environment Diagram.....	11
Figure 7. BioIntelligence Center (BIC) Model.....	13
Figure 8. Statistical Anomalies Module page (demonstration data).....	18
Figure 9. Hospital Chief Complaints and Diagnosis page indicating elevated Gastrointestinal count a (demonstration data)	18
Figure 10. Time Series Module page (demonstration data)	19
Figure 11. Patient List page (demonstration data)	20
Figure12. Patient Map page (demonstration data)	20
Figure 13. Patient Map page options (demonstration data)	21
Figure 14. Describe page (demonstration data)	21

Tables

Table 1. Top Metropolitan Statistical Areas (by population).....	4
Table 2. Sample W2 Count and W2 Rate Calculations(Weekday)(demonstration data)..	14



A. Introduction

BioSense Background

BioSense is the national public health surveillance program hosted by the Centers for Disease Control and Prevention (CDC). It is designed to improve the nation's capability for disease control and monitoring through access to real-time healthcare data from healthcare organizations, state syndromic surveillance systems, national laboratories, and other data sources. Data are analyzed and made available to public health at local, state, and federal levels for conducting biosurveillance and health situational awareness.

A BioSense user can employ the data available in BioSense to:

- Monitor the size, location, and rate of spread of an outbreak.
- Provide local, state, and nationwide [health situational awareness](#) for suspect illnesses before, during and after an outbreak.
- Monitor seasonal trends of influenza and other disease indicators
- Assist in case-finding for epidemiologic investigations

Specifically, BioSense focuses on:

- Data Transmission: assuring the secure, timely, and routine receipt of health data for public health surveillance.
- Data Analysis: using statistical methods to detect events and enabling public health to use these methods to make informed decisions.
- 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 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.

BioSense Data Sources

BioSense can be scaled to fit the capacity and needs of local users and has the ability to interoperate with other data systems. Healthcare data from large hospital systems, data aggregating systems, and state/county surveillance systems are linked to BioSense providing a unified national view and creating a “system of systems.”

Data sources include:

- State/regional surveillance systems
- Private hospitals and hospital systems

- National laboratories
- Department of Defense (DoD) and Department of Veterans Affairs (VA) healthcare outpatient medical facilities
- Outpatient pharmacies
- Sentinel provider and WHO laboratory data from the Influenza Division, CDC

BioSense participating facilities comprise over 50 major metropolitan areas in the USA including Los Angeles, Chicago, Washington DC, and Philadelphia (Table 1). BioSense receives real-time data from over 450 clinical care hospitals, which provide a broad national view and local data perspective. Over 1100 VA and DoD hospitals and ambulatory care facilities participate in BioSense, further strengthening surveillance coverage in the United States (Figure 1). This interoperable, unified system allows for large scale data management and customizing of data queries.

Table 1: Top Metropolitan Statistical Areas (by population)

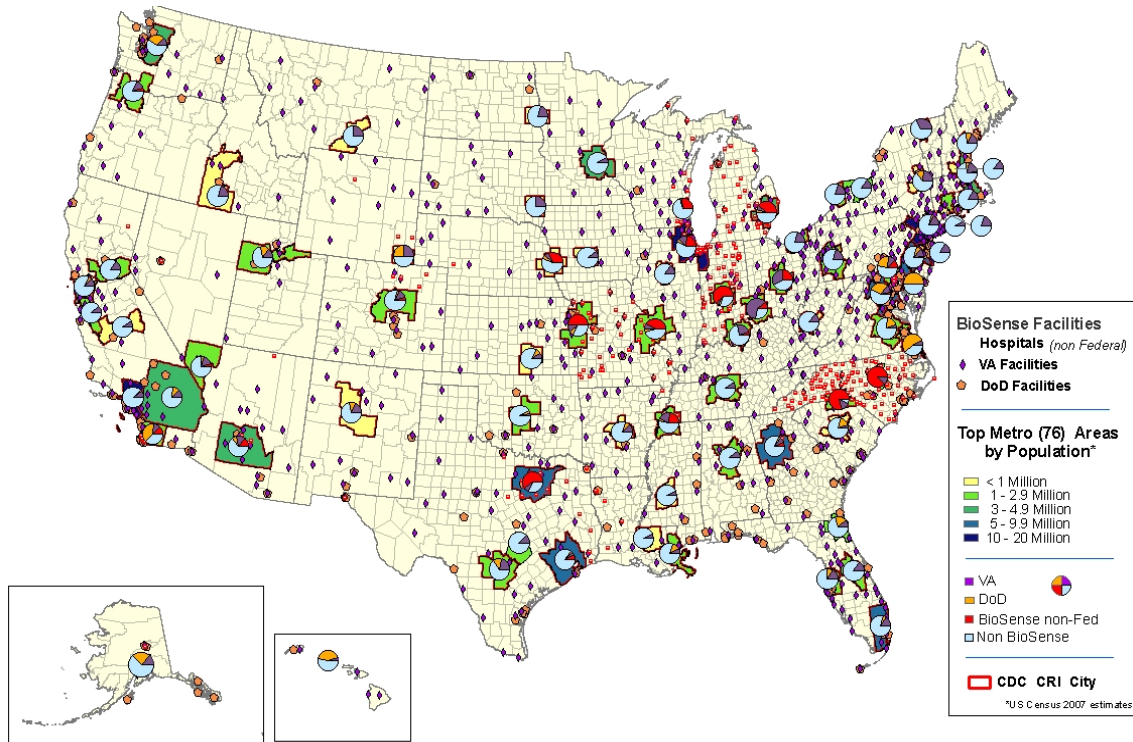
Rank (1-76)	Metropolitan Statistical Area (MSAs)	US Census Estimate 2007	Hospitals not participating in BioSense*	Hospitals participating in BioSense*	DoD Facilities	VA Facilities
	Total	176,684,040	2,608	224	168	300
1	New York-Newark-Edison, NY-NJ-PA**	18,815,988	249	0	6	34
2	Los Angeles-Long Beach-Santa Ana, CA**	12,875,587	162	1	3	14
3	Chicago-Naperville-Joliet, IL-IN-WI**	9,524,673	99	26	5	14
4	Dallas-Fort Worth-Arlington, TX**	6,145,037	60	45	1	6
5	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD**	5,827,962	117	1	5	10
6	Houston-Baytown-Sugar Land, TX**	5,628,101	98	5	1	3
7	Miami-Fort Lauderdale-Miami Beach, FL	5,413,212	79	0	3	11
8	Washington-Arlington-Alexandria, DC-VA-MD-WV**	5,306,565	57	2	21	7
9	Atlanta-Sandy Springs-Marietta, GA**	5,278,904	68	3	2	4
10	Boston-Cambridge-Quincy, MA-NH**	4,482,857	96	2	4	14
11	Detroit-Warren-Livonia, MI**	4,467,592	36	17	2	3
12	San Francisco-Oakland-Fremont, CA**	4,203,898	63	0	2	8
13	Phoenix-Mesa-Scottsdale, AZ**	4,179,427	46	6	3	6
14	Riverside-San Bernardino-Ontario, CA**	4,081,371	47	0	5	6
15	Seattle-Tacoma-Bellevue, WA**	3,309,347	25	0	7	3
16	Minneapolis-St. Paul-Bloomington, MN-WI**	3,208,212	40	0	0	3
17	San Diego-Carlsbad-San Marcos, CA**	2,974,859	28	6	19	5
18	St. Louis, MO-IL**	2,803,707	37	24	2	5

Rank (1-76)	Metropolitan Statistical Area (MSAs)	US Census Estimate 2007	Hospitals not participating in BioSense*	Hospitals participating in BioSense*	DoD Facilities	VA Facilities
19	Tampa-St. Petersburg-Clearwater, FL**	2,723,949	46	0	5	7
20	Baltimore-Towson, MD**	2,668,056	31	4	8	4
21	Denver-Aurora, CO**	2,464,866	34	1	1	3
22	Pittsburgh, PA**	2,355,712	57	0	0	9
23	Portland-Vancouver-Beaverton, OR-WA**	2,175,113	23	1	0	3
24	Cincinnati-Middletown, OH-KY-IN**	2,133,678	14	1	0	7
25	Cleveland-Elyria-Mentor, OH**	2,096,471	32	0	0	5
26	Sacramento--Arden-Arcade--Roseville, CA**	2,091,120	21	0	0	3
27	Orlando, FL**	2,032,496	25	0	0	4
28	San Antonio, TX**	1,990,675	42	0	5	8
29	Kansas City, MO-KS**	1,985,429	26	20	3	5
30	Las Vegas-Paradise, NV**	1,836,333	21	1	1	3
31	San Jose-Sunnyvale-Santa Clara, CA**	1,803,643	18	0	0	2
32	Columbus, OH**	1,754,337	14	3	0	4
33	Indianapolis, IN**	1,695,037	18	19	1	1
34	Virginia Beach-Norfolk-Newport News, VA-NC**	1,658,754	23	0	16	2
35	Charlotte-Gastonia-Concord, NC-SC**	1,651,568	6	11	0	2
36	Providence-New Bedford-Fall River, RI-MA**	1,600,856	30	0	0	3
37	Austin-Round Rock, TX	1,598,161	31	0	0	2
38	Milwaukee-Waukesha-West Allis, WI**	1,544,398	31	6	0	1
39	Nashville-Davidson--Murfreesboro, TN**	1,521,437	50	0	0	3
40	Jacksonville, FL	1,300,823	19	0	3	2
41	Memphis, TN-MS-AR**	1,280,533	26	6	1	5
42	Louisville, KY-IN**	1,233,735	30	1	0	6
43	Richmond, VA**	1,212,977	23	0	3	1
44	Oklahoma City, OK**	1,192,989	41	0	1	1
45	Hartford-West Hartford-East Hartford, CT**	1,189,113	22	0	0	1
46	Buffalo-Cheektowaga-Tonawanda, NY**	1,128,183	23	0	0	4

Rank (1-76)	Metropolitan Statistical Area (MSAs)	US Census Estimate 2007	Hospitals not participating in BioSense*	Hospitals participating in BioSense*	DoD Facilities	VA Facilities
47	Birmingham-Hoover, AL**	1,108,210	29	0	0	2
48	Salt Lake City, UT**	1,099,973	15	0	2	2
49	Raleigh-Cary, NC	1,047,629	4	7	0	1
50	Rochester, NY	1,030,495	14	0	0	2
51	New Orleans-Metairie-Kenner, LA**	1,030,363	48	0	4	3
54	Honolulu, HI**	905,601	18	0	9	1
55	Fresno, CA**	899,348	14	0	0	1
57	Albany-Schenectady-Troy, NY**	853,358	18	0	2	4
58	New Haven-Milford, CT**	845,494	13	0	0	2
60	Albuquerque, NM**	835,120	22	0	1	2
61	Omaha-Council Bluffs, NE-IA**	829,890	16	5	1	1
67	Baton Rouge, LA**	770,037	23	0	0	1
69	Columbia, SC**	716,030	23	0	3	1
78	Little Rock-North Little Rock, AR**	666,401	24	0	1	2
84	Wichita, KS**	596,452	18	0	1	1
86	Boise City-Nampa, ID**	587,689	11	0	0	2
91	Des Moines, IA**	546,599	12	0	0	1
93	Jackson, MS**	534,047	20	0	0	1
98	Portland-South Portland, ME**	513,102	11	0	1	2
124	Manchester-Nashua, NH**	402,302	7	0	0	1
133	Peoria, IL**	371,206	6	0	0	1
136	Trenton-Ewing, NJ**	365,449	13	0	0	1
137	Anchorage, AK**	362,340	7	0	2	1
151	Charleston, WV**	303,950	11	0	0	1
191	Sioux Falls, SD**	227,171	11	0	0	2
197	Burlington-South Burlington, VT**	207,361	2	0	0	1
214	Fargo, ND-MN**	192,417	7	0	0	1
262	Dover, DE**	152,255	1	0	1	0
266	Billings, MT**	149,657	3	0	0	1
421	Cheyenne, WY**	86,353	3	0	1	1

* Includes non-federal hospitals/ ** Cities Readiness Initiative (CRI) city

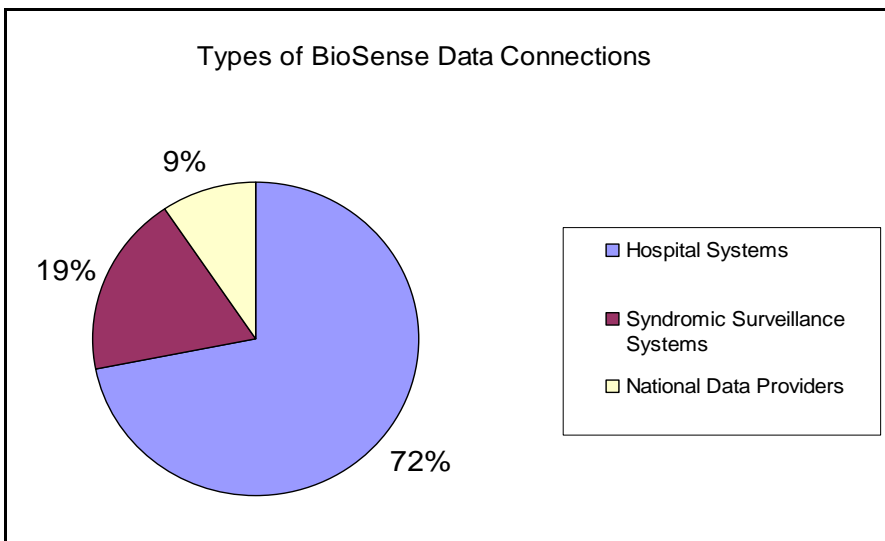
Figure 1. BioSense Participation in Top 76 MSAs as of June 2008



Hospital data sources by type of connection to BioSense consist of 23 healthcare organizations (private hospitals or hospital systems), 6 syndromic surveillance systems, 1141 VA and DoD facilities, and 3 national laboratories (Figure 2).

As of May 2008, hospital data from 26 states is received by BioSense. Within those states, a total of 454 hospitals are sending data to BioSense (Table 2).

Figure 2. BioSense Data Connections



BioSense Data Types

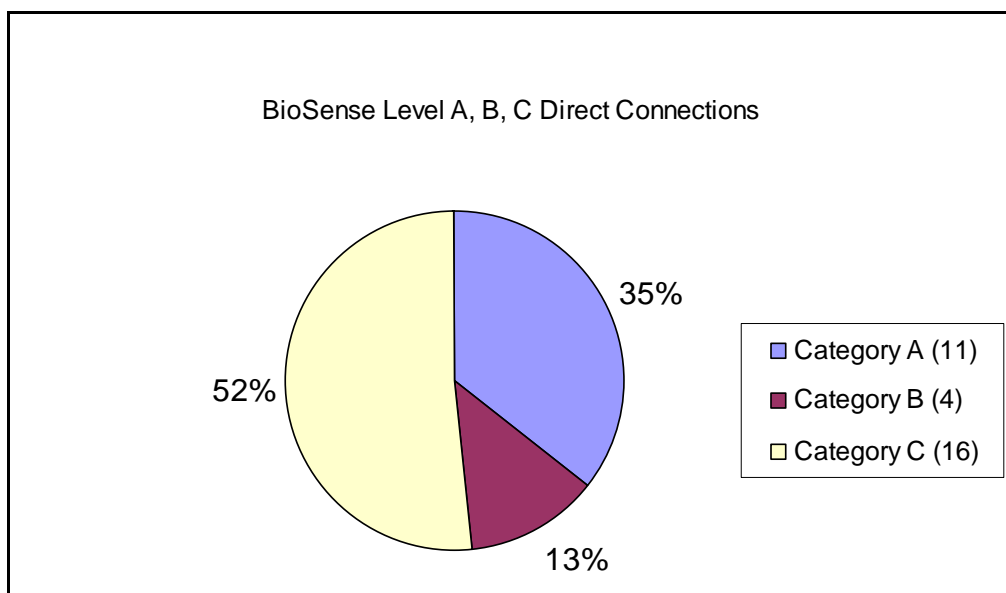
BioSense receives and analyzes the following data types:

- **Foundational Data**— anonymized demographics, physician diagnoses, patient chief complaints, and daily hospital census
- **ED Clinical Data**— emergency department data, vital signs, triage notes, and discharge summaries
- **Laboratory Data**— microbiology laboratory orders and results
- **Pharmacy Data**— medication orders
- **Radiology Data**— radiology laboratory orders and results

Data sources are categorized as levels A, B, or C. Category A data is defined as foundational data which includes chief complaints, demographics, working diagnosis, discharge diagnosis, disposition, hospital utilization, and clinical data (vitals, triage notes, discharge summary). Category B data is Category A plus laboratory data (microbiology results). Category C data is Category A and B plus pharmacy orders and radiology orders or interpretation results. BioSense receives data from 11 Category A's, 4 Category B's, and 16 Category C's (Figure 3).

Most hospitals send foundational and ED clinical data to BioSense. VA and DoD hospital facilities send [ICD-9 CM](#) diagnosis code to BioSense, which includes Current Procedural Terminology (CPT) codes and patient demographics, zip code, and disposition. Laboratory data is a national batched data feed that consists of Laboratory Corporation of America, Quest Diagnostic, and RelayHealth (formerly Per-Se)/ NDC Health data. This data includes ICD-9 CM, diagnosis test orders, patient demographics and zip codes.

Figure 3. BioSense Hospital Level A, B, C Direct Connections

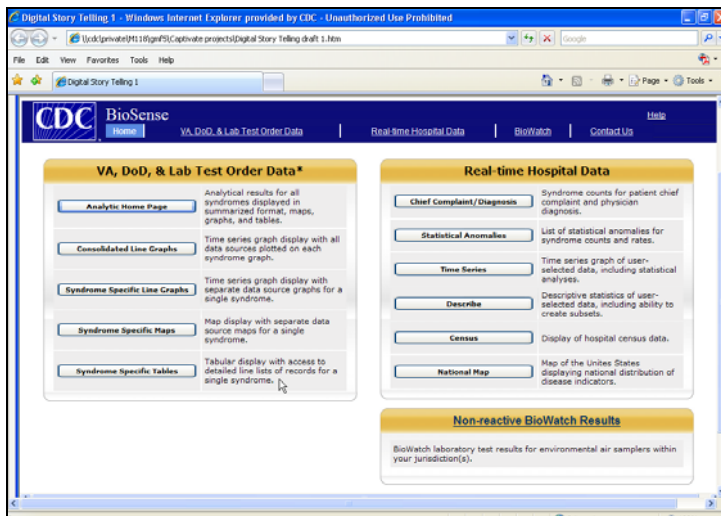


Access to BioSense

BioSense is used by local, state, and federal public health officials and hospital staff. Users include individuals such as epidemiologists, bioterrorism response coordinators, and hospital infection control personnel. While personnel from a given facility may view only data from that facility, state or metropolitan reporting area (MRA), officials may view data for all facilities in their jurisdiction.

Each public health jurisdiction or facility/facility group has a BioSense Administrator who is responsible for authorizing access to BioSense for appropriate staff within their jurisdiction or hospital. To obtain contact information for the appropriate BioSense Administrator for your jurisdiction, send an e-mail request to BioSensehelp@cdc.gov. Upon approval from the local BioSense administrator, a user must apply for a [digital certificate](#) and be granted access. Once granted access, the user may then log on to BioSense via CDC's [Secure Data Network \(SDN\)](#).

Figure 4. BioSense homepage screen image



B. How BioSense Functions

PHIN Architecture

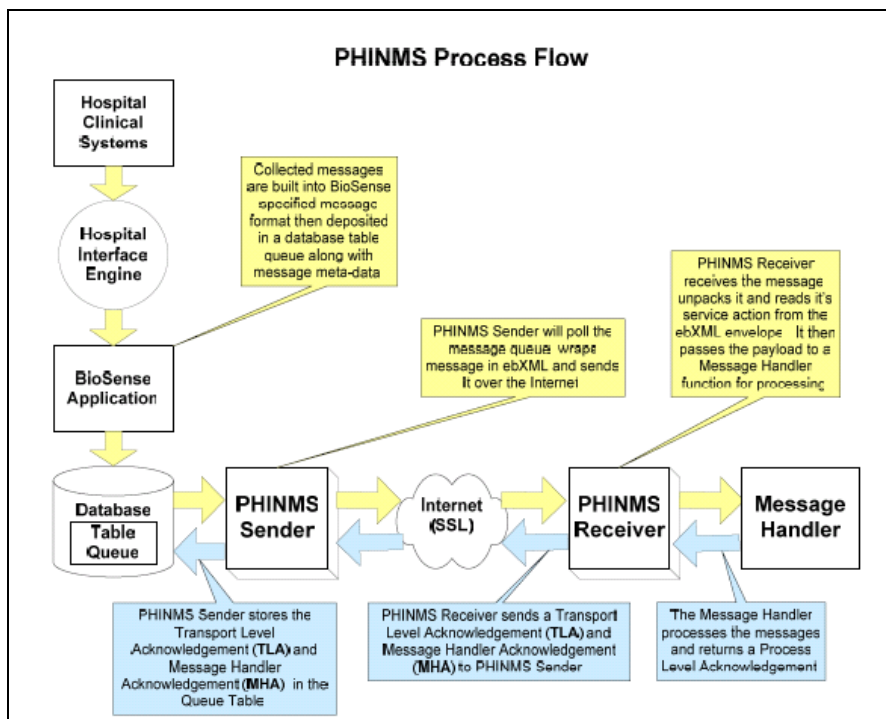
BioSense requires data sources to transmit data using the Public Health Information Network Messaging System (PHINMS). PHINMS is an interoperable messaging system developed by CDC for data sources to use in sending private data securely over the Internet in batches (transmitted once per day) or real-time, streaming data. PHINMS communications consist of standardized messages and vocabulary that make it possible for public health organizations to communicate with each other.

PHINMS is a JAVA based application consisting of PHINMS sender (client) and PHINMS Receiver (Server). The PHINMS Sender is deployed on a Java Application

Server running on a server or a workstation at the data source’s site. The PHINMS Receiver resides within the Secure Data Network (SDN) at CDC. The data is securely exchanged over the Internet using [Secure Sockets Layer \(SSL\)](#). In a typical BioSense data connection with a data source, the data source’s site extracts the data of interest to BioSense from data source clinical systems through an interface engine. The extracted data are processed and transformed into [HL7](#) messages conforming to CDC’s specifications. The messages are deposited into a database table queue where they are stored in the PHINMS Sender and processed to be transported over the Internet to the PHINMS Receiver (Figure 5). Health department data, connected via a syndromic surveillance system, comprises the majority of the data that is received from BioSense. Hospitals/hospital systems have one to a few dozen facilities where data is received from BioSense (Table 2).

The PHINMS Receiver obtains the messages, checks the signature, retrieves the HL7 messages, then drops the HL7 message into a query for processing to determine what kind of message it is. Once processed, an acknowledgement is sent back to the PHINMS Sender.

Figure 5. BioSense Technical Process Flow Diagram



BioSense Operating Environment

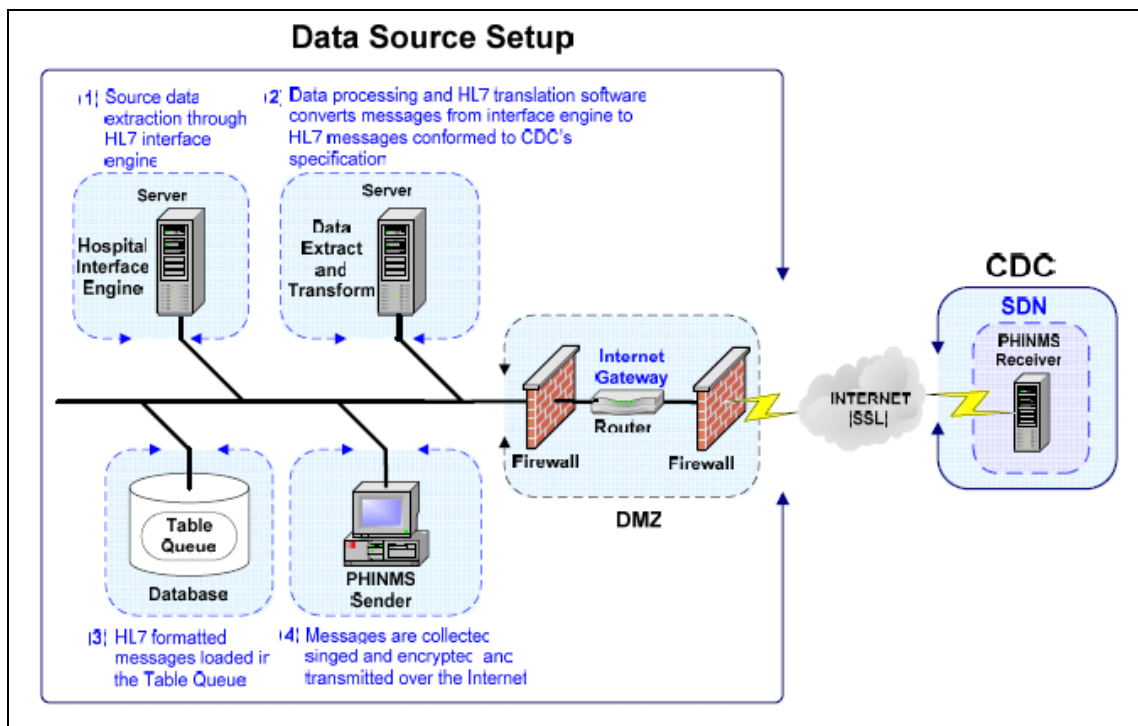
Basic equipment requirements for the data source set up environment are:

- “Always ON” Internet connection with firewall protection.

- Windows/UNIX/Linux Server or workstation capable of running the PHINMS software, Tomcat Application Server (Bundled with PHINMS), and Java – JDK1.4 (Bundled with PHINMS) or higher.
- BioSense data processing and HL7 translation software. Typical functions of the software include:
 - BioSense ID management
 - Data filtering and de-identification
 - Standard vocabulary mapping
 - HL7 message transformation
- Server capable of running BioSense data processing and HL7 translation software.
- Server capable of running JDBC compliant Database Server.

Figure 6 demonstrates how PHINMS software can be integrated into the data source's network and securely connected to the Internet:

Figure 6. BioSense Operating Environment Diagram



C. Data Processing

Due to the challenging nature of electronic data submission and maintenance of surveillance systems, ensuring the validity of data received for conducting surveillance

and monitoring health can be difficult. Challenges of the multifaceted process of capturing electronic data include:

- Exchange of large quantities of data, many data types, and multiple disease indicators
- Integrating multiple systems' data feeds into the application
- Errors in reporting clinical data
- Lack of standardized data analysis strategies
- Difficulty of interpreting data from limited data sets

Data Quality

Hospital facilities and state/local syndromic systems that transmit data to CDC every 15 minutes can view BioSense data in real time. However, state/local syndromic systems may transmit some of the information received from their respective data sources only once per day (batch data).

The latency period to view data types in the BioSense application varies with each data source. Typically, VA or DoD ICD -9 CM codes take 3-7 days to view patient visits. Hospital clinical information such as chief complaints are transmitted within 24 to 48 hours. Final diagnoses take about 7 days and hospital census data is received daily by CDC.

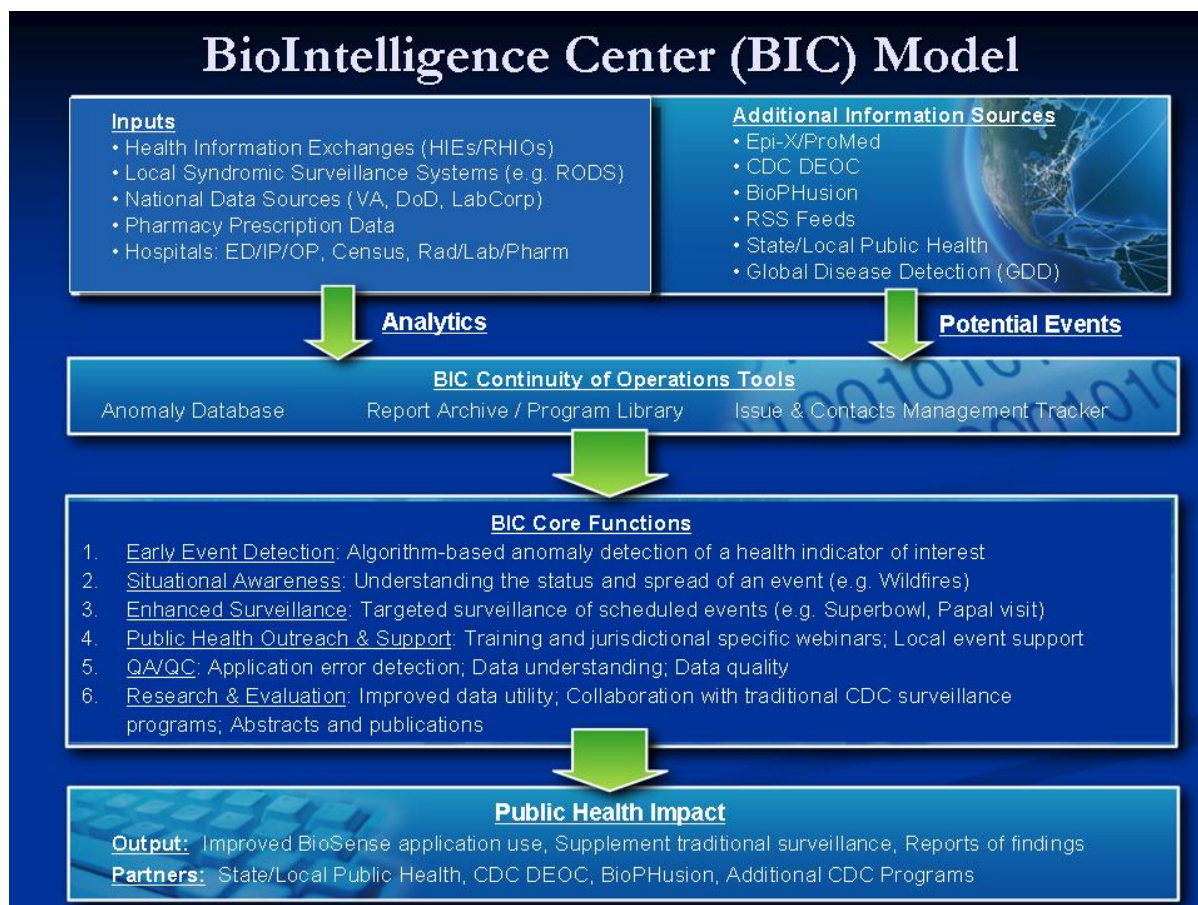
As the data flows through the data source via the PHINMS receiver to the BioSense application, they go through various levels of data quality analysis. The data are modified and processed several times at specific data transfer points as they travel from the data source to the BioSense application and to CDC. Throughout this process, data quality analysis teams within CDC are assigned to create analytical data subsets, coordinate data quality reviews, and develop BioSense data quality reports to identify unusual data outliers and to validate data calculations.

Data Analysis Steps

Analysts in the CDC BioIntelligence Center (BIC), are trained in epidemiology, public health, biostatistics, informatics, and emergency response. Each day they 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.

BIC analysts use web-based media reports and public health communications sources to identify potential public health events and to target monitoring efforts. An established set of standardized tools are used to assist in the six core functions established by the BIC: Early Event Detection, Situational Awareness, Enhanced Surveillance, Public Health Outreach and Support, Quality Assurance and Control, and Research and Evaluation (Figure 7).

Figure 7. BioIntelligence Center (BIC) Model



BIC analysts interpret BioSense data and detect data anomalies using the BioSense application’s methods. These methods include W2count and W2rate. Recent improvements in the W2 statistical algorithms, which improve accuracy of recognizing true data anomalies, are scheduled to be implemented in Fall 2008. Data anomaly detection involves calculating the “Expected” Value, and then comparing it with the “Observed” value, i.e. [Residual = Observed – Expected], test statistic = Residual divided by Standard deviation, followed by calculating a recurrence interval (analogous to a p-value). The following steps describe how the BioSense Application calculates and analyzes hospital and VA/DoD data.

- Calculate the Expected Value

For hospital data anomaly detection, syndrome and sub-syndrome counts are analyzed by calculating the Expected Value using a modified EARS C2 approach, termed “W2”. This method indicates recent data changes through comparison with a 7-day moving average. W2 is an Expected Value (7-day moving average) separated into weekday and weekends/holidays (with a 2-day lag); data for weekends/holidays are compared to a 7-day moving average including only weekends/holidays (with a 2-day lag).

- Calculate the Residual

= the difference between the Observed Value and the Expected Value.

- Calculate the Test Statistic (number of standard deviations above or below the Expected Value)

=residual/standard deviation

- Calculate the Recurrence Interval

This method uses the frequency distribution of the Test Statistic (i.e., residual divided by standard deviation). The top 1% value is equivalent to an empiric “p-value” of approximately 0.01 and a recurrence interval of 100 days. Recurrence intervals are based on historical data and will be periodically calculated. BioSense users can set a recurrence interval threshold for the Statistical Anomalies module. Current thresholds (in days) that a user may select are 50, 100, 500, and 1000 days.

Table 2. Sample W2 Count and W2 Rate Calculations (Weekday)

Note: Today is a Wednesday 12/15 and the current syndrome count is 94.

Date	Day of Week	Visits	Count	Include	Expected Value W2Count	Expected Value W2Rate	Residual W2Rate
12/02	Thursday	329	71	Yes	82.7	81.2	10.2
12/03	Friday	327	54	Yes	85.7	79.5	25.5
12/04	Saturday	275	75	No (weekend)	70.8	64.5	10.5
12/05	Sunday	313	71	No (weekend)	70.8	73.4	2.4
12/06	Monday	381	86	Yes	74.9	84.4	1.6
12/07	Tuesday	332	73	Yes	74.9	73.6	0.6
12/08	Wednesday	325	74	Yes	74.9	72	2.0
12/09	Thursday	304	72	Yes	75.9	66.5	5.5
12/10	Friday	315	64	Yes	74.6	67.3	3.3
12/11	Saturday	310	64	No (weekend)	71.5	74.1	10.1
12/12	Sunday	240	64	No (weekend)	71.5	57.4	6.6
12/13	Monday	252	61	No (lag day)	70.6	53.8	7.2
12/14	Tuesday	435	117	No (lag day)	70.6	92.9	24.1
12/15	Wednesday	394	94		70.6	84.1	9.9

W2Count

- Expected value= mean of (71,54,86,73,74,72,64) = 70.6
- Standard deviation = std dev of (71,54,86,73,74,72,64) = 9.8
- Residual = 94-70.6 = 23.4
- W2c = residual/standard deviation = 23.4/9.8 = 2.4

W2Rate

- N = sum of (71,54,86,73,74,72,64) = 494

- $D = \text{sum of } (329,327,381,332,325,304,315) = 2313$
- $\text{Expected value} = (N/D) * \text{visits} = (494/2313) * 394 = 84.1$
- $\text{Estimated standard deviation} = \text{mean of } (10.2, 25.5, 1.6, 0.6, 2.0, 5.5, 3.3) = 6.9$
- $\text{Residual} = \text{observed} - \text{expected} = 94 - 84.1 = 9.9$
- $W2r = \text{residual} / \text{est standard deviation} = 9.9 / 6.9 = 1.4$

W2Count

- $\text{Expected value} = \text{mean of } (69,67,60,46,60,51,63) = 59.4$
- $\text{Standard deviation} = \text{std dev of } (69,67,60,46,60,51,63) = 8.3$
- $\text{Residual} = 81 - 59.4 = 21.6$
- $W2c = \text{residual} / \text{standard deviation} = 21.6 / 8.3 = 2.6$

W2Rate

- $N = \text{sum of } (69,67,60,46,60,51,63) = 416$
- $D = \text{sum of } (295,297,329,247,313,311,281) = 2073$
- $\text{Expected value} = (N/D) * \text{visits} = (416/2073) * 335 = 67.2$
- $\text{Estimated standard deviation} = \text{mean of } (7.8,5.7,7.9,6.7,6.7,11.6,6.4) = 7.5$
- $\text{Residual} = \text{observed} - \text{expected} = 81 - 67.2 = 13.8$
- $W2r = \text{residual} / \text{est standard deviation} = 13.8 / 7.5 = 1.8$

For VA/DoD data anomaly detection, the SMART Model is used, but will be supplemented with the W2 method in the near future. The SMART model requires more than 1 year of baseline data. Individual models run for each syndrome-zip code-data source-day. The SMART model includes parameters for day of week, holidays, day after holidays, and seasonality with an intercept difference for each zip code to calculate the Expected Value.

BIC analysts complete daily baseline reports that provide knowledge regarding expected data flow for each facility. This involves comparing counts displayed in today's "Chief Complaint and Diagnosis" module to previous days. The Chief Complaint/Diagnosis is a quick overview of syndrome activity by day for patient chief complaint and diagnosis data stratified by patient class, indicating statistically significant syndrome counts.

BIC analysts check daily reports provided by application development, data warehouse, and data provisioning to ensure non-duplicative reporting. To ensure non-duplicative data or readmissions of the same patient in the same day, a Patient ID and Analysis Visit ID are created. The Patient ID contains no personally identifiable information. Instead, it contains a de-identified number for the patient that can be used by trained facility staff to re-identify a patient (based on state and local public health department investigation). The Analysis Visit ID is a random number for the BioSense definition of a patient visit. It combines all "visits" occurring within a 24 hour period.

Characterization of Data Anomalies

The BIC analysts identify anomalies by automated statistical testing for syndrome and sub-syndrome count and rate. Anomalies are based only on real-time hospital data that is

received from the application. The anomaly dataset is de-duplicated, filtered, and entered into a Microsoft Access database (the [Anomaly Database](#)). Anomalies are prioritized and characterized based on a set of standard criteria. The BIC analysts portray data anomalies by positive, descriptive, and negative criteria.

Data Pre-Processing

Clinical condition classifiers (“syndromes” and “sub-syndromes”) are used for biosurveillance. Diagnostic and pre-diagnostic data is collected in electronic form obtained from [ICD-9 CM codes](#), chief complaint free text, and others. The data is then aggregated into groups (“bins”) such as syndromes and sub-syndromes.

Data anomalies are examined within 11 syndromes, which are designed to capture symptoms associated with critical bioterrorism agents. Syndromes include:

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

BioSense data are further parsed into 78 sub-syndromes, allowing for a more in-depth view of the data. Sub-syndromes allow BioSense users to monitor trends for infectious, chronic, injury, and other disease indicators. Many but not all sub-syndromes map directly to a syndrome. For example:

- Botulism-like
 - Paralysis
 - Speech disturbance
 - Dysphagia
- Gastrointestinal
 - Abdominal pain
 - Anorexia
 - Diarrhea

Detailed syndrome definitions for diseases associated with critical bioterrorism-associated agents and a list of ICD-9-CM codes included in each syndrome are available at: <http://www.bt.cdc.gov/surveillance/syndromedef/index.asp>

Mapping of ICD-9 CM codes to 78 sub-syndromes were developed by BioSense medical expert staff. The sub-syndromes allow a more granular view of events than using the 11 syndromes mentioned above.

D. Making Decisions

How is the data used by the CDC?

The BIC accesses the BioSense application to identify and describe potential disease outbreaks or bioterrorism events, thereby providing a national health situational awareness view. If the BIC identifies a data anomaly that is deemed to be of potential public health significance, the analyst shares a report with the BIC team. If CDC leadership concurs, the analyst contacts the state and/or local health department using jurisdictionally-defined protocols, provides the summary of relevant findings, and supplies any needed analytic support to follow up on the public health investigation. Reports are also provided to CDC DEOC during activation of the DEOC.

How is the data used by state and local public health and hospital facilities?

Approved state and local public health departments can access BioSense data to monitor and track suspect illnesses, before, during, and after an outbreak or bioterrorism attack. Participating hospital facilities can also access the data to help identify and describe the size, location, and rate of an outbreak within their facility. Access and customizable options to view existing patient data provides important insight into health events and trends within a state and local public health jurisdiction and a healthcare facility.

E. Application Outputs

The BioSense application provides analytical support to the DEOC and to state/local public health during catastrophic events. Actual events during which BioSense has supplied such support have included:

- California wildfires
- Oregon/Washington State border flooding
- Tennessee tornados

Figures 8–14 present a scenario from the BioSense application for monitoring a local community’s public health.

1. An epidemiologist from a public health department first looks at the Statistical Anomalies module during a specific time period to see elevated observed counts for disease indicators (Figure 8).

Figure 8. Statistical Anomalies Module page

Syndrome	Data Type	Patient Class	Anomaly Date	Anomaly Location	Observed Count	Count Recurrence Interval (Days)	Expected Count*	Obs./Exp.	Rate per 1000 Visit	Rate Recurrence Interval (Days)
Rash	Final Diagnosis	Emergency Department	11/29/2007		3	50	0.2	21.0	120.0	50
Respiratory	Final Diagnosis	Inpatient	11/29/2007		6	40	2.6	2.3	500.0	100
Gastrointestinal	Final Diagnosis	Emergency Department	11/27/2007		7	100	2.6	2.7	269.2	200
Gastrointestinal	Final Diagnosis	Emergency Department	11/27/2007		5	100	1.2	4.3	416.6	90
Respiratory	Chief Complaint	Emergency Department	11/27/2007		4	60	0.9	4.6	800.0	80
Respiratory	Chief Complaint	Emergency Department	11/26/2007	Hospital 18	5	100	0.9	5.8	625.0	100
Gastrointestinal	Working Diagnosis	Emergency Department	11/25/2007		6	40	2.6	2.3	333.3	80

- The epidemiologist decides to monitor the Gastrointestinal (GI) syndrome closely due to a recent increase in *community gastrointestinal disease* outbreak near his community. The epidemiologist accesses the Chief Complaint page to focus on occurrences within the GI syndrome. The BioSense application flagged unusual GI-related activity on a specific day (Figure 9).

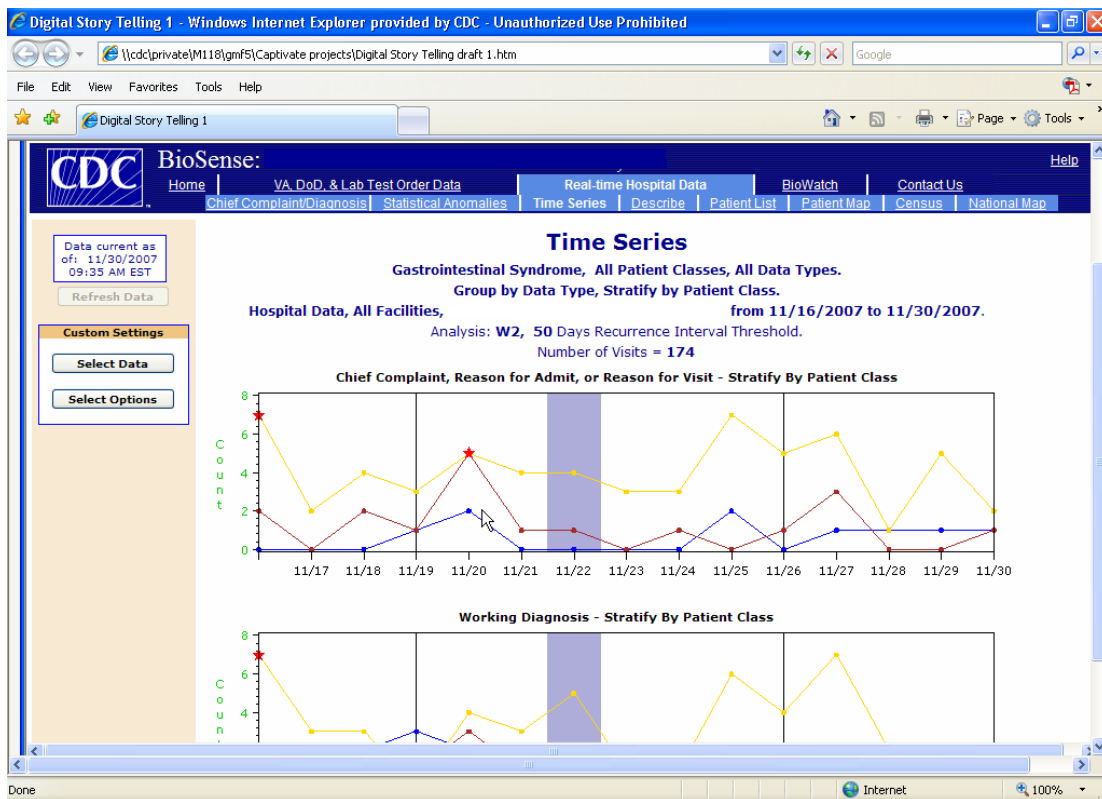
Figure 9. Hospital Chief Complaints and Diagnosis page indicating elevated Gastrointestinal count

Syndrome	Reason for Visit	Patient Class								
		Outpatient			Emergency Department			Inpatient		
		Working Diagnosis	Final Diagnosis	Chief Complaint	Working Diagnosis	Final Diagnosis	Reason for Admit	Working Diagnosis	Final Diagnosis	
Botulism-like	0	0	0	0	0	0	1	0	0	
Fever	0	0	0	1	1	0	0	0	2	
Gastrointestinal	0	0	0	7	7	1	2	1	2	
Hemorrhagic Illness	0	0	0	1	0	0	1	0	1	
Localized Cutaneous Lesion	0	1	0	0	0	2	1	0	1	
Lymphadenitis	0	0	0	0	0	0	0	0	0	

- The epidemiologist is taken to the time series page when he/she clicks on the highlighted number of gastrointestinal (GI) patients. The star on the graph depicts a statistically significant result. This graph allows the epidemiologist to evaluate trends over time for specific syndromes and sub-syndromes and to identify patterns. The epidemiologist can customize specific parameters for the data he/she wants to view such as looking at trends in *community gastrointestinal disease* related sub-syndromes.

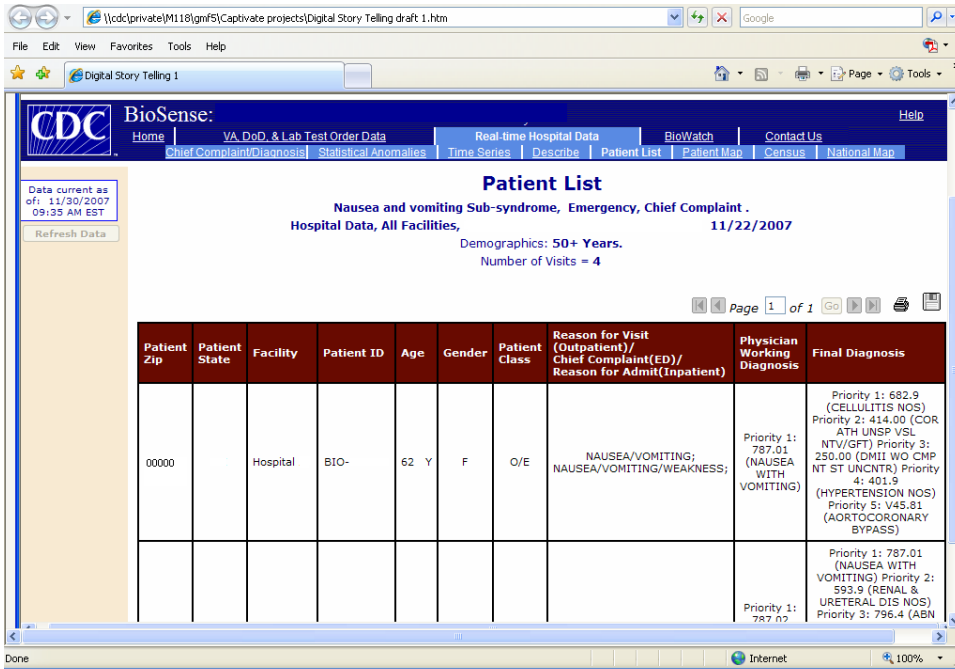
The time series page also displays three graphs, one per data type (chief complaints, working diagnosis, and final diagnosis), each of which is stratified by patient class (outpatient, emergency department, and inpatients) (Figure 10).

Figure 10. Time Series Module page



- The epidemiologist is also able to navigate to a patient list or a map by clicking on the star in the Time Series graph or the Patient List link for the day associated with the GI statistical anomaly. From the Patient List, the epidemiologist can view the locations, demographics, chief complaints, and diagnoses of the patients associated with the anomaly. The epidemiologist is able to view patient level data for the patients in the GI cluster. He/she is able to save Patient list data tables to Excel by clicking on the “floppy disk” icon on the application screen (Figure 11).

Figure 11. Patient List page



- The epidemiologist can click on the Patient Map link, which displays syndrome data by patient residence zip code. The map also offers options to display state, county, or zip code boundaries, in addition to interstate highways (Figures 12 & 13).

Figure 12. Patient Map page

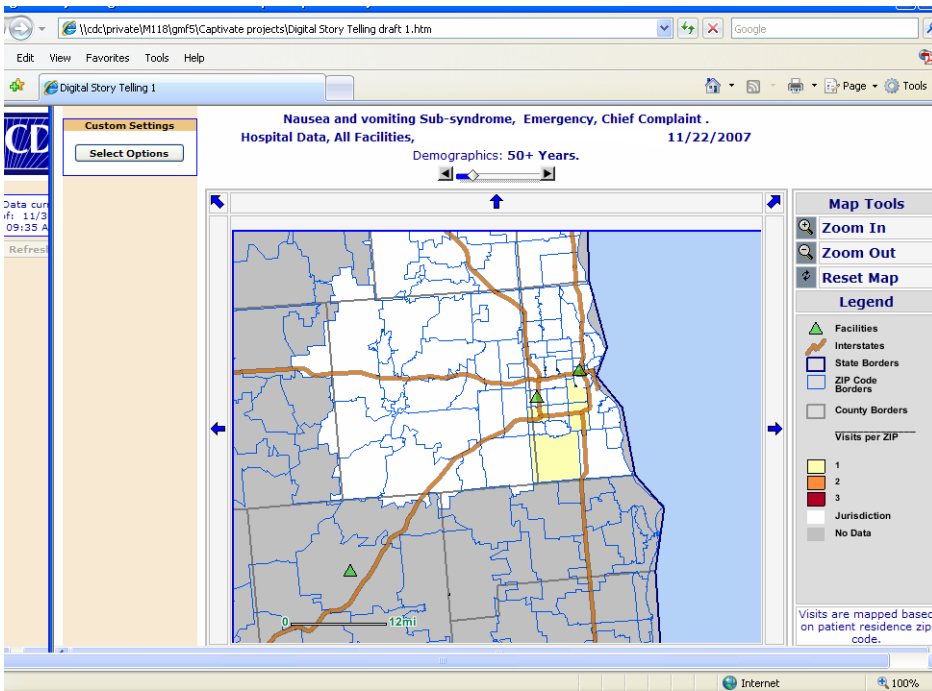
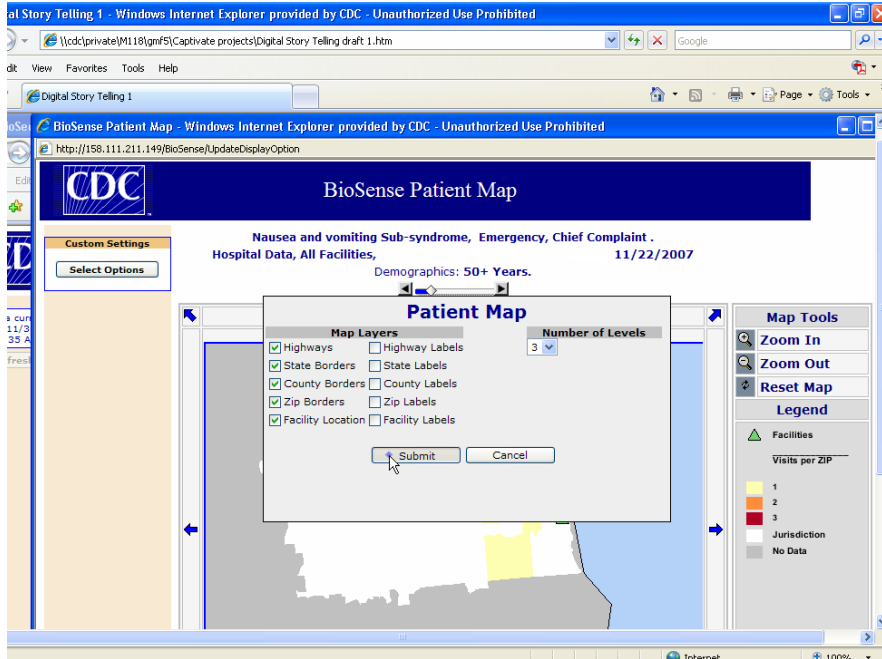
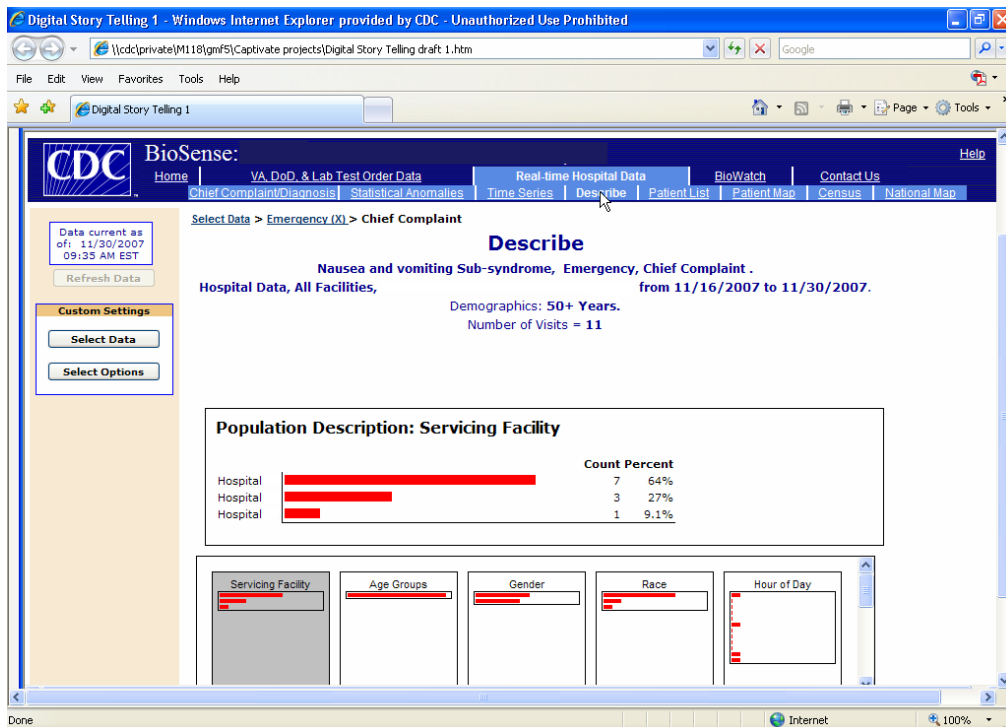


Figure 13. Patient Map page options



6. The epidemiologist can click on the link for the Describe page to examine descriptive statistics for age group, race, and hospital facility and to examine the sub-syndrome distributions for the patients in the cluster of interest (Figure 14).

Figure 14. Describe Page



In summary, the epidemiologist logged onto the BioSense application and spent about 30 minutes reviewing data in his community. As a result, he was able to understand and analyze the *community gastrointestinal disease* outbreak better. He shared this information with the local public health department, giving them the opportunity to prepare in advance and make informed decisions for protecting the health of their communities.

F. For More Information

The BIC analysts monitor, analyze, and interpret BioSense data at a national level daily. They also offer technical assistance to BioSense users via e-mail at BioSensehelp@cdc.gov. Through the BIC, BioSense users can request help with the BioSense application, receive additional information about the program, or provide feedback to help improve the BioSense program.

For questions about the secure data network or BioSense technical support, contact the Technical Help Desk: 1-800-532-9929.

To sign-up with the BioSense Listserv and receive the BioSense Bulletin and BioSense related news and events, e-mail BioSenseusers@cdc.gov.

Glossary

Admissions, Discharge, Transfer (ADT) – Message type of clinical health data used in Health Level 7 formatted communication.

American Health Information Community (AHIC) – Federal advisory organization that makes recommendations to the Secretary of the U.S. Department of Health and Human Services (HHS) on how to accelerate the development and adoption of health information technology standards, certification, and policies on a national level.

Anomaly Database – Location of an anomaly identified by statistical algorithms.

CDC Director’s Emergency Operation Center (DEOC) – Focal point for all decision making for all incoming and outgoing disaster information at the CDC.

Current Procedure Terminology (CPT) codes – Listing of descriptive terms and identifying codes for reporting medical services and procedures as they are performed by physicians.

Digital Certificate – Certificate that is downloaded onto your computer allowing you to view secure applications. It is obtained by applying to the CDC Secure Data Network.

Epi-X– CDC’s internet based system for the exchange of information among public health about public health threats. Through Epi-X, CDC staff, state and local public health departments, poison control centers, and other public health officials can access and share preliminary outbreak information quickly and securely.

Health Level 7 (HL7) – Commonly accepted standard for the electronic exchange of clinical health data.

Health Situational Awareness – Ability to use detailed, real-time health data to confirm or refute the existence of an outbreak and monitor its size, geography, rate of change, and life cycle in support of an effective outbreak response.

International Classification of Diseases 9th Revision (ICD-9-CM codes) – A numerical list of the disease code numbers to classify diseases and a wide variety of signs and symptoms. The ICD is used world-wide for morbidity and mortality statistics and decision support in medicine

Real-time – Delivery of healthcare data to BioSense from hospitals, laboratories, ambulatory care settings, and other health data sources 15 minutes to 24 hours after being received. Latency time varies by data type.

Secure Data Network (SDN) – Provides security controls to applications like BioSense that exchange data between CDC and public health partners. External partners who access the applications within the SDN are required to have digital certificates installed on their computers to verify their identity when they log into the system.

Secure Socket Layer (SSL) – Protocol for delivering private messages via the Internet. It uses a cryptographic system that uses two keys to encrypt data—a public key known to everyone and a private or secret key known only to the recipient of the message.