RODS: A Real-time Public Health Surveillance System

Three articles detailing structure and usage.

Technical Description of RODS

- "Technical Description of RODS: A Real-time Public Health Surveillance System." Tsui, Espino, Dato, Gesteland, Hutman, and Wagner. JAMIA 10:5 (2003).
- Background: U Pitt, etc

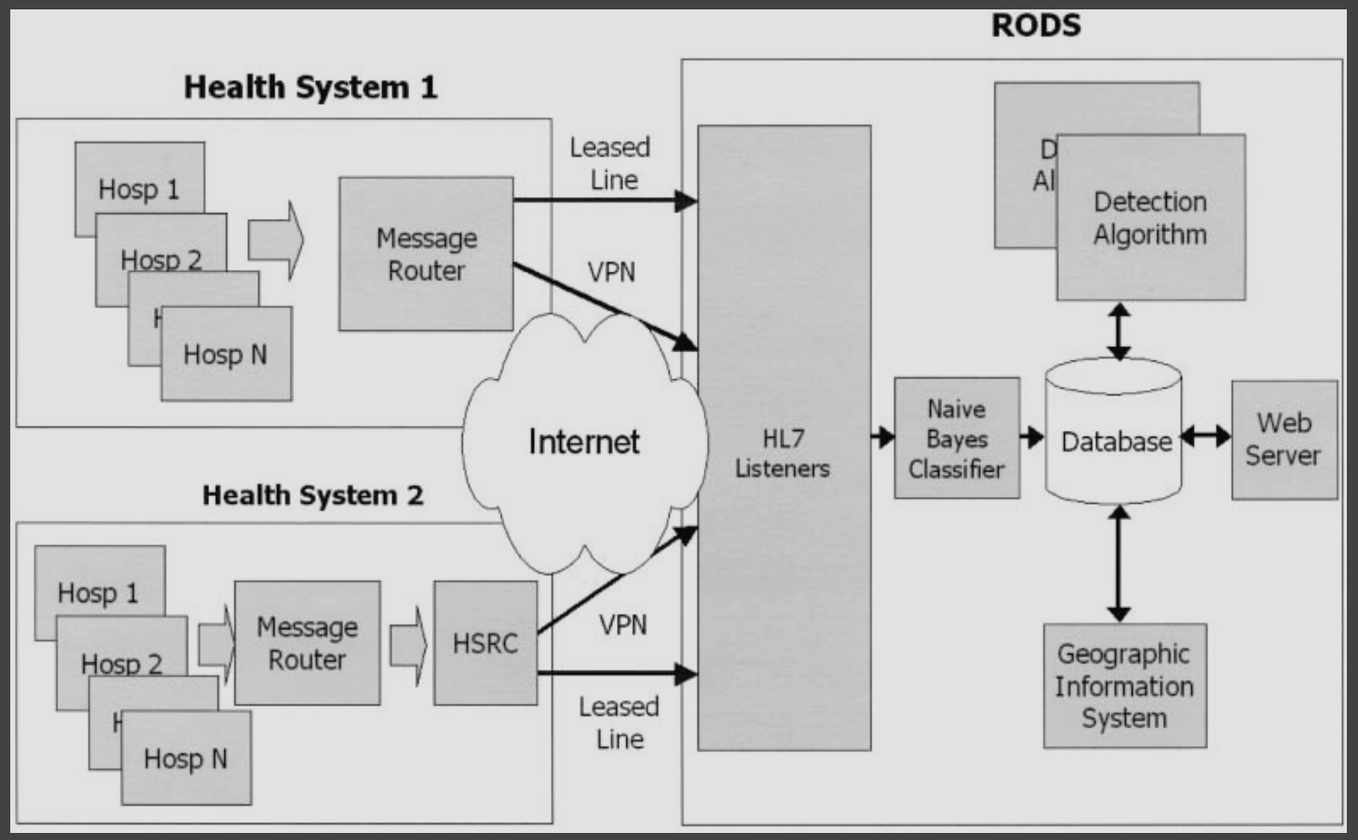
Background

- Public Health Surveillance has become increasingly important while we fight the “War Against Terror.”
- 2001 syndromic surveillance solution is slower (delayed transfer of data 24 hours or more) and/or more costly (staffed round-the-clock).
- Current systems do not use communication standards and protocols.

Why RODS?

- Standards compliance.
- Near real-time data aggregation through HL7.
- Relatively cheap cost.

Basic RODS Structure



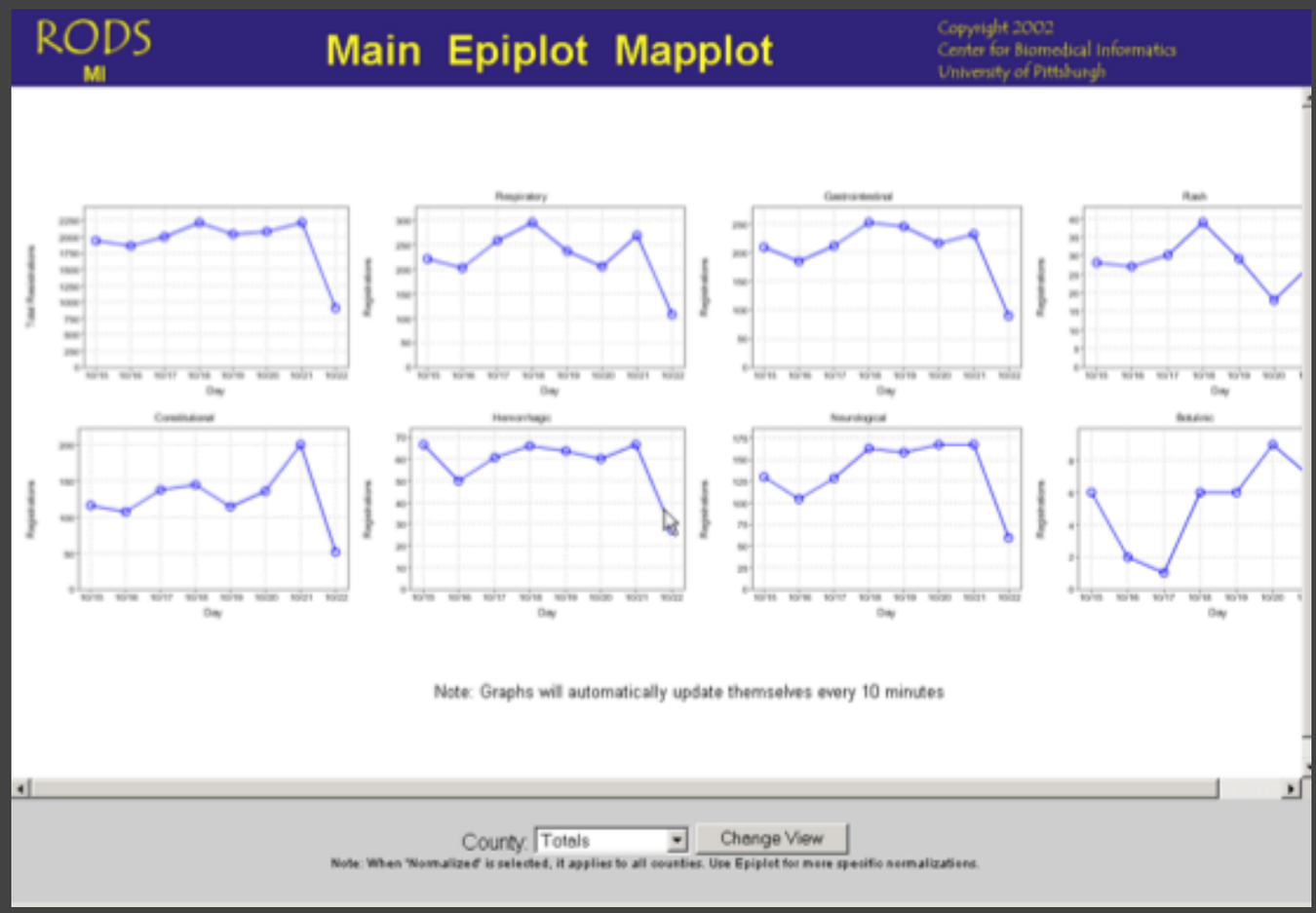
RODS Data

- Shared among multiple health institutions.
- Data type is HL7 (discussion later).
- Transmission (chief complaint free text via HL7 messaging servers).
- Integrity (avoid incoming duplicates via composite primary key).
- Database (Oracle 8i).

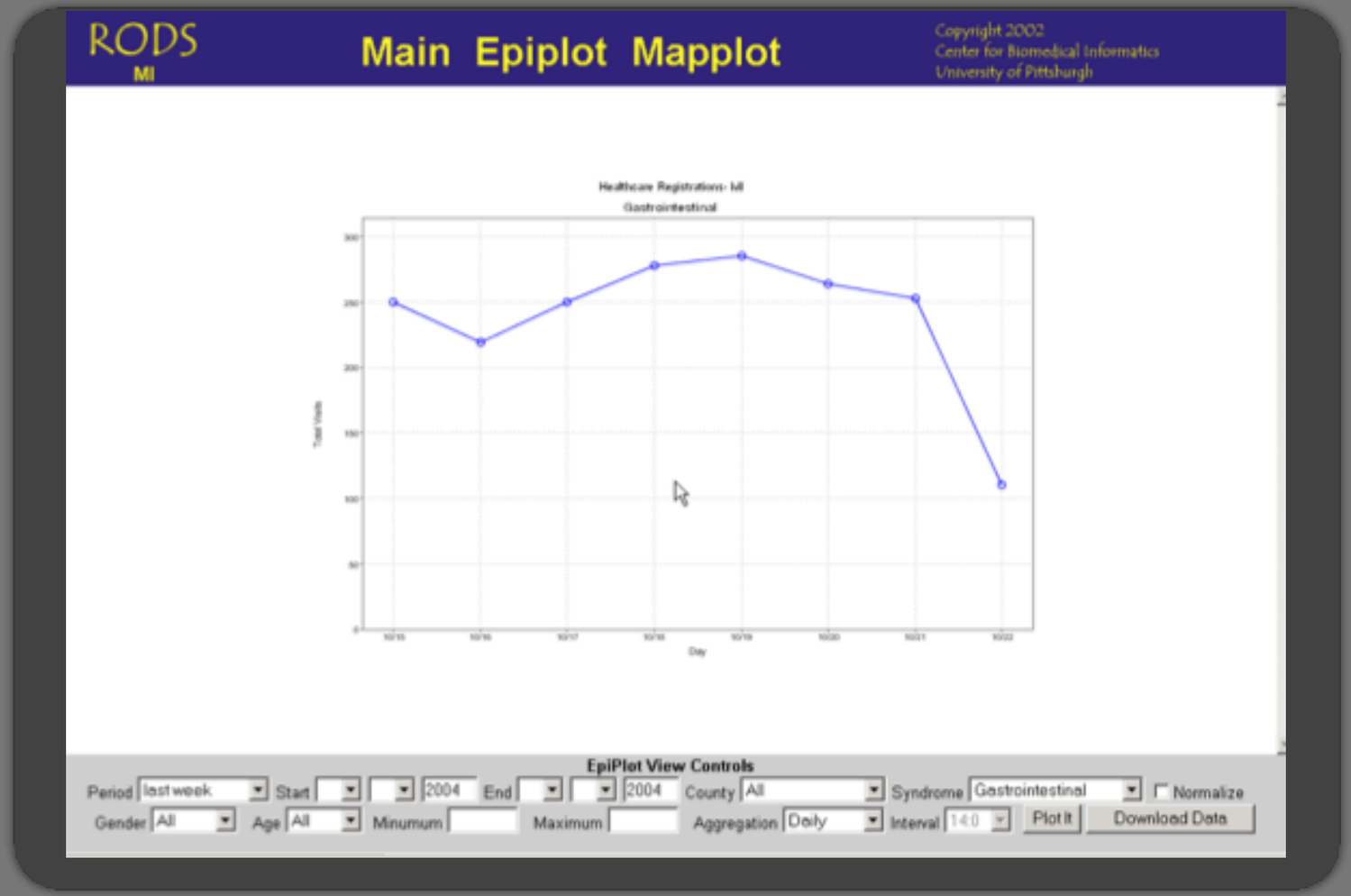
RODS: Application Level

- Natural Language Processing:
  Bayesian Complaint Coder (CoCo)
- Detection Algorithms (recursive least square)
  http://en.wikipedia.org/wiki/Recursive_least_squares_filter
- Alert Notification
- User Interface

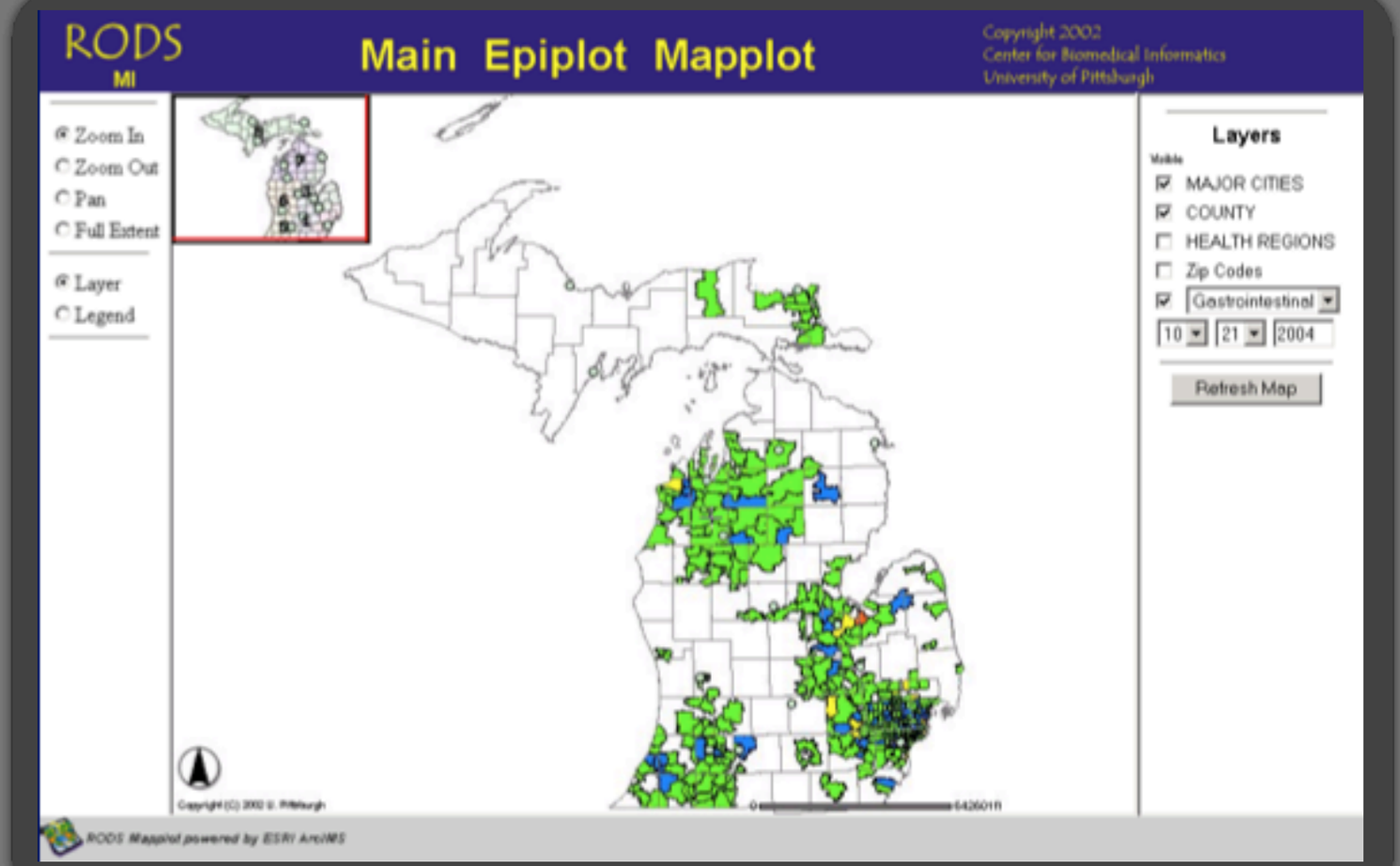
RODS: User Interface



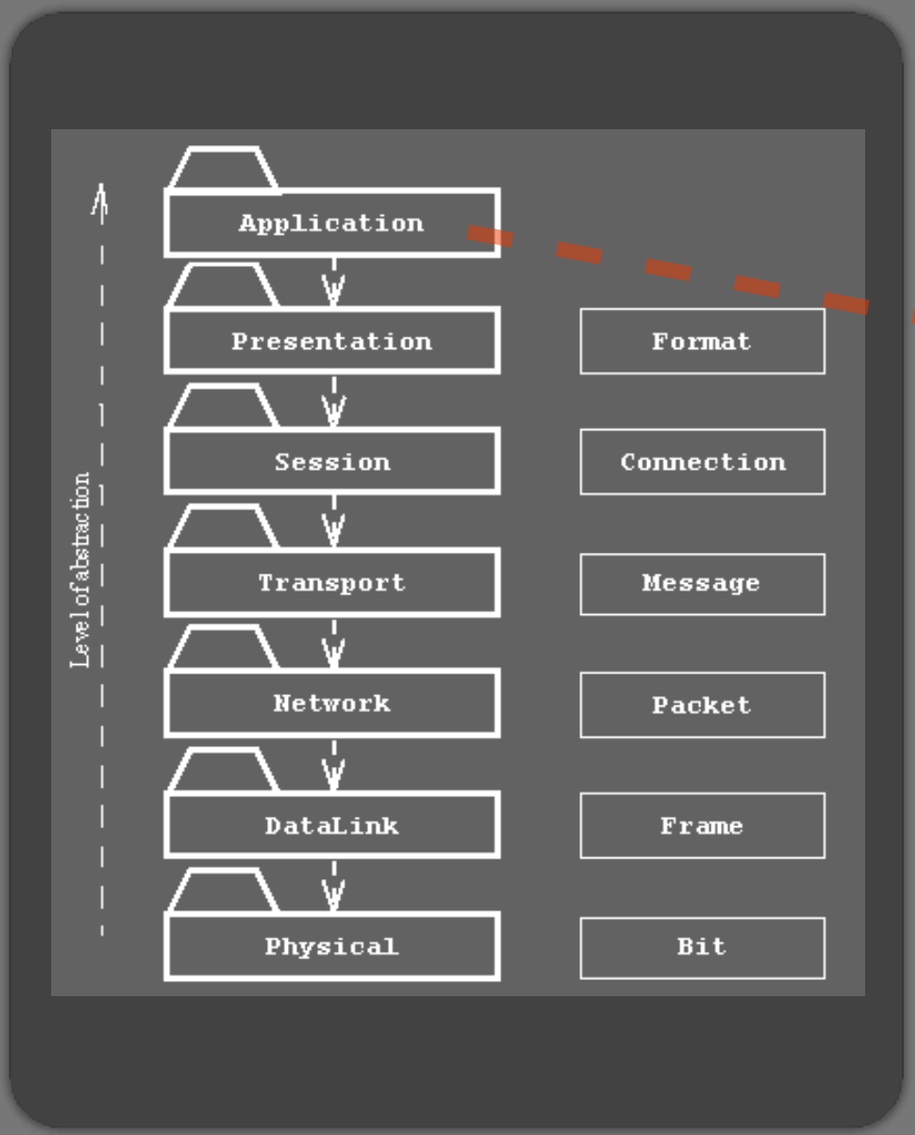
RODS: User Interface



RODS: User Interface



HL7 (Health Level 7) Messaging



- An ANSI standard for healthcare specific data exchange between computer applications. The name comes from "Health Level 7", which refers to the top layer (Level 7) of the Open Systems Interconnection (OSI) layer protocol for the health environment.

- http://www.hl7.org/
  http://www.hl7.ca/

HL7 (Health Level 7) Messaging

```
MSH|^~\&|EPIC|EPICADT|SMS|SMSADT|199912271408|CHARRIS|ADT^A04|1817457|D|2.3|
EVN|A04|199912271408|||CHARRIS
PID||0493575^^^2^ID 1|454721||DOE^JOHN^^^^|DOE^JOHN^^^^|19480203|M||B|254 E238ST^^EUCLID^OH^44123^USA||(216)731-4359|||M|NON|400003403~1129086|999-|
NK1||CONROY^MARI^^^^|SPO||(216)731-4359||EC|||||||||||||||||||||||||
PV1||O|168   ~219~C~PMA^^^^^^^^^^|||277^ALLEN FADZL^BONNIE^^^^|||||||||  ||2688684|||||||||||||||||||||199912271408||||||002376853
```

- **MSH segment** contains information about the Sender and Receiver of the message, the type of the message, a time stamp, etc.
- **EVN** contains information about the type of message; for example, A04 (Register a patient). *Information contained in EVN is duplicated in MSH, so starting from HL7 version 2.3 this segment is excluded from all message definitions.*
- **PID** contains demographic information about the patient such as name, id codes, address, and so on.
- **PV1** contains information regarding the patient's stay in the hospital, such as location assigned, referring doctor, etc.
- HL7 in XML? http://www.interfaceware.com/manual/ch-6-3.html

National Electronic Disease Surveillance System (NEDSS)

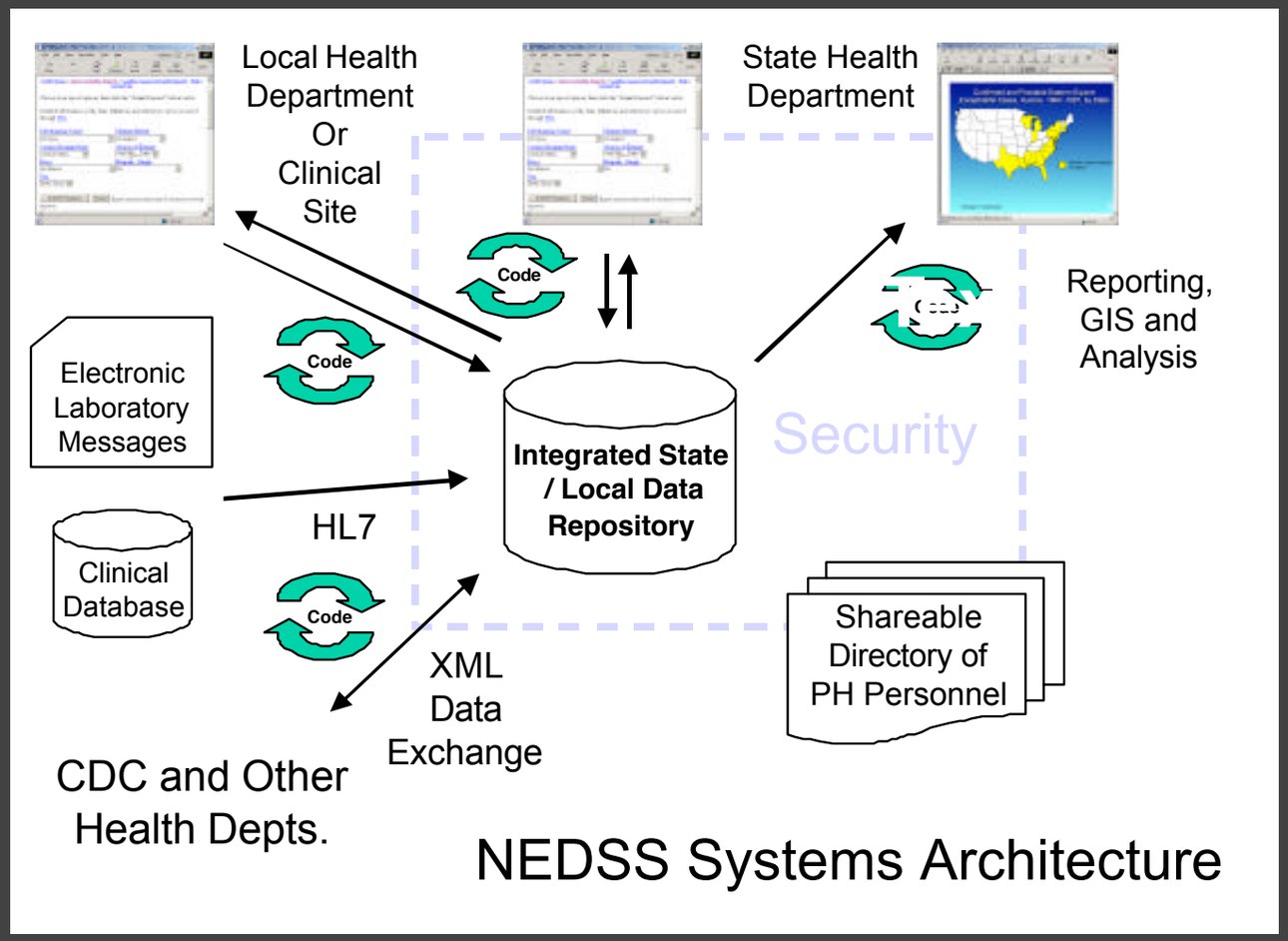
- The National Electronic Disease Surveillance System (NEDSS) is an initiative that promotes the use of data and information system standards to advance the development of efficient, integrated, and interoperable surveillance systems at federal, state and local levels. A primary goal of NEDSS is the ongoing, automatic capture and analysis of data that are already available electronically.

- NEDSS is based on the following principles:
  - Utilization of industry standards
  - Reliance on off-the-shelf software
  - Internet-based secure transmission of data
  - A common "look and feel" of systems
  - Common reporting requirements
  - No requirement to use specific software

- NEDSS is designed to address the limitations of current surveillance systems which include:
  - Multiple incompatible disease specific systems that currently exist
  - Incomplete and delayed data
  - Burden on health care system to report disease
  - Overwhelming volume of data to be managed by health departments
  - Lack of state-of-the-art information technology

- The mission is to design and implement seamless surveillance and information systems that take advantage of the best information and surveillance technology, and serve the following needs at the local, state, and national levels:
  - Monitor and assess disease trends
  - Guide prevention and intervention programs
  - Inform public health policy and policy makers
  - Identify issues needing public health research
  - Provide information for community and program planning
  - Protect confidentiality while providing information to those who need to know

National Electronic Disease Surveillance System (NEDSS)



http://www.cdc.gov/nedss/

Conclusions (Round 1)

- Real-time monitoring objectives not quite met but close.
- Identified two data types (free-text chief complaint information and OTC sales) which have 70% coverage.
- Importance of HL7 routers and adherence to NEDSS architecture.
- Giving it away via GNU licensing is generally insufficient. Expertise requirements are too high. Application service provider model much better.
- Need for computing component behind firewall to act as a case detector in a distributed environment.

What's Next?

- Develop early-warning capability for a large, outdoor release of anthrax, by means of improved interfaces, more mature detection algorithms to reduce false alarms, and the highly specialized, distributed look-back function.
- Enlarge the application service provider to include more states and data types.
- Add additional disease scenarios (e.g., in-building anthrax release, vector-borne and food-borne diseases, SARS).

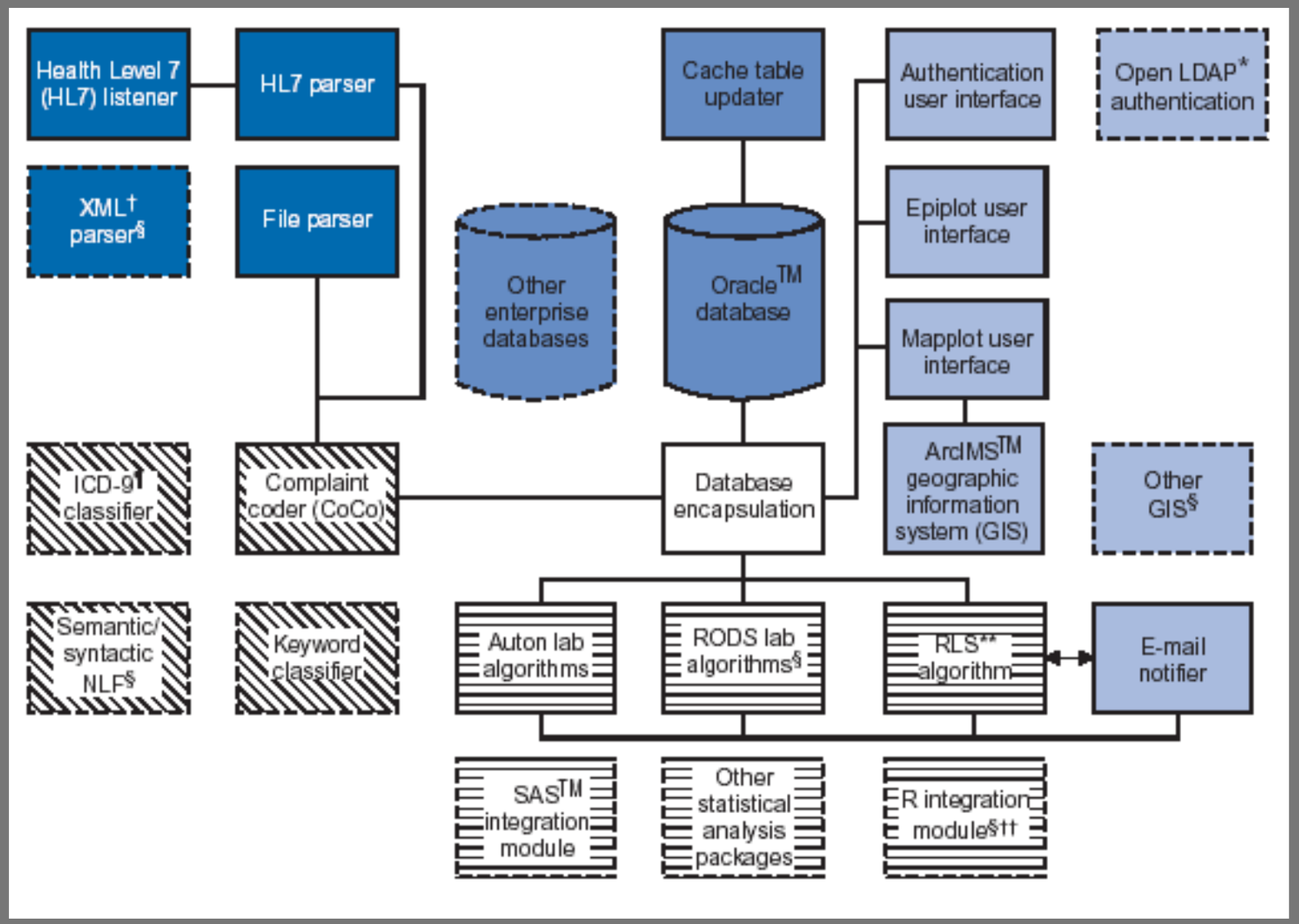
RODS 2: Electric Boogaloo ↘

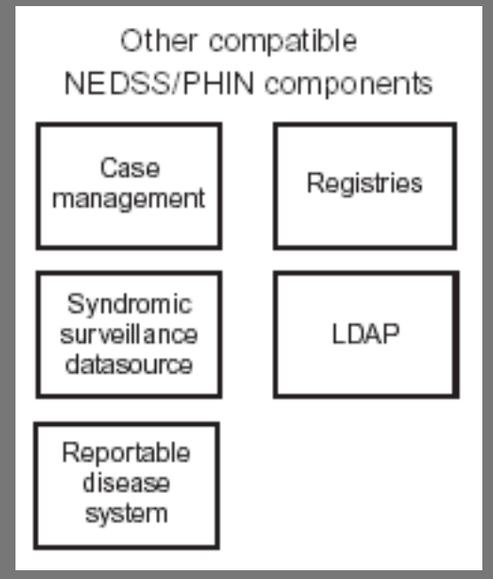
- The creation of OpenRODS (http://openrods.sourceforge.net/) addresses a need to allow a broad user base to develop and customize RODS to their needs. Following general open-source philosophy, this licensing, by creating more active and targeted development, should benefit the entire RODS community.
- New features in 2.0: drill-down of age and sex, customized jurisdictions, a simplified GIS interface, and user preferences.

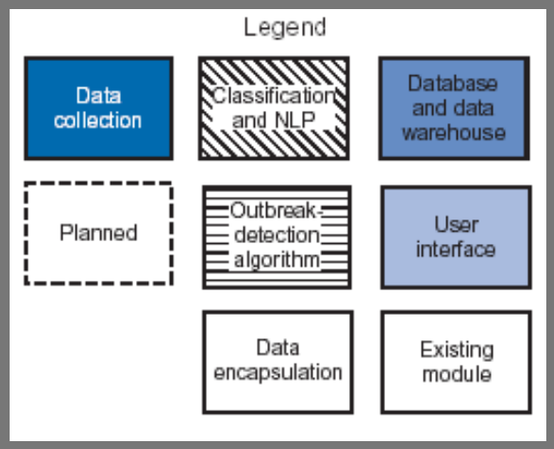
RODS 2.0 Architecture

- Six functional areas:
  - data collection (HL7 systems, text-parsers, possible XML component?)
  - syndrome classification (Coco using Naive Bayesian classifier to assign chief complaint data to user-defined categories; intended release of ICD-9-based classification)
  - data warehousing
  - database encapsulation (EJBs for accessing database)
  - outbreak detection (recursive least squares)
  - user interface (authentication, display of data, GIS interface)

RODS 2.0 Architecture







Current Status of RODS

TABLE 1. Existing features of the Real-Time Outbreak Disease Surveillance (RODS) system, version 2.0, and features awaiting development

RODS feature	Exists in RODS 2.0	Exists as GPL*-compatible source code	Needs to be developed or tested
Data-collection modules			
Health Level 7 (HL7) listener	X		
HL7 parser for microbiology reports			X
HL7 parser for admissions, discharge, and transfer (ADT) messages	X		
Text file parser	X		
XML parser		X	
Syndrome-classification modules			
Simple Bayesian syndrome classifier	X		
Syntactic/semantic natural language processing (NLP) classifier		X	
Keyword classifier			X
ICD-9† classifier			X
Multiple data-type classifier		X	
Data-warehousing modules			
Diverse database options		X	
Integrated data-warehouse engine			
Aggregation by sex and age	X		
Outbreak-detection modules			
Integrates with external statistical analysis tools	X	X	X
Recursive least-squared (RLS) detection algorithm	X		
User-interface modules			
Manual data-entry interface		X	
Diverse geographic information system software options		X	
Lightweight directory access protocol (LDAP) interface		X	
Time-series graphing	X	X	
Options and preferences	X		
Custom jurisdictions	X		
E-mail notifier	X		
Database-encapsulation modules			
Database encapsulation	X		

* GNU General Public License.
† *International Classification of Diseases, Ninth Revision.*

RODS in Seven Weeks: Salt Lake City, 2002 Olympics

- The events of September 11, 2001, along with the subsequent domestic anthrax terrorism incident, highlighted the need for a rapid-response surveillance system at the 2002 Winter Olympic Games in Salt Lake City.
- RODS was used in order to test the real-world rapid deployment of a syndromic surveillance system.

Background

- An example of common practice in syndromic surveillance is the CDC's "drop-in" method, which consists of manual collection of data by health clinicians.
- But while "drop-in" surveillance is more thorough, it is also more costly, more obtrusive, and ultimately slower to respond. It is the obtrusive nature of the "drop-in" method that led the Utah Dept of Health to choose an alternative.

Objectives

- Adequate coverage of the Utah region.
- Chief complaint data returned in a “timely fashion.”
- Analysis of data for aberrant patterns.
- Notification of appropriate officials.
- Facilitation of investigation of incident.

Surveillance

- Daily polling of select physicians, pharmacies, vets, poison control, forensic pathologists.
- 24-hour surveillance of urgent care facilities.
- Manual syndromic surveillance system reviewing logs.
- RODS used to monitor ~2M people in Wasatch Front region. Tracking was focused on geographical location, not participants themselves.
- RODS collaborated with public health officials and health systems (70% of regional coverage).

Operation

- Main data type collected: free-text chief complaint.
- HL7 data transferred via VPN back to the RODS servers at the University of Pittsburgh, parsed by *Coco*.
- Data then analyzed every four hours by the automated detection algorithm, recursive least square (RLS) adaptive algorithm.
  - RLS was chosen because it required only a few days to generate its model coefficients.
  - Also RLS is more sensitive to recent historical data to predict outcomes, making it ideal for a short-term event like the Olympics.

Operation

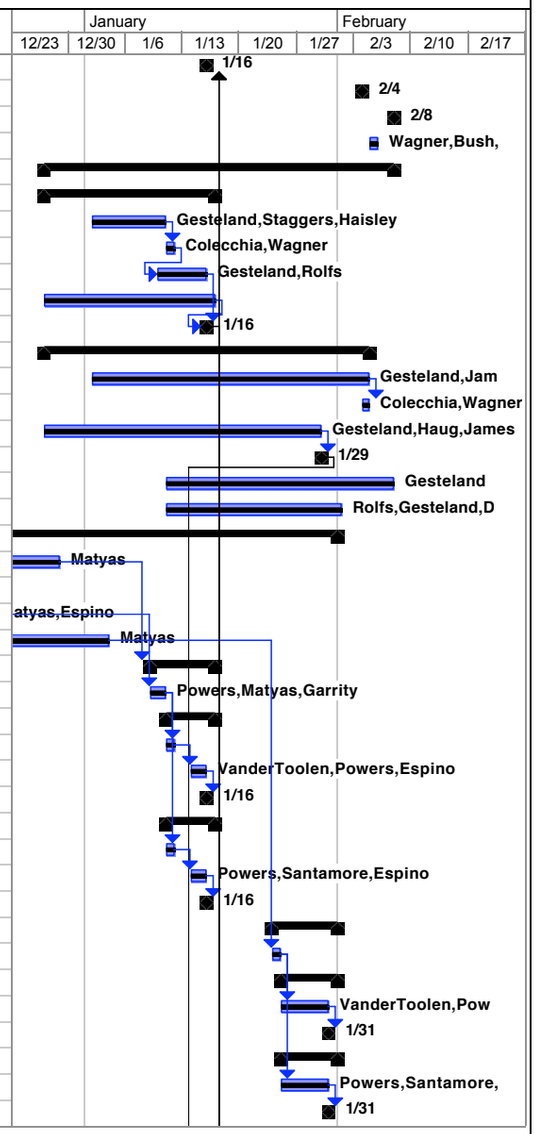
- What's Strange About Recent Events (WSARE) was also used to analyze the incoming data for anomalous events. But because it requires more data to normalize, it was not used in the formal response structure at the Olympics.
- RODS was set to alert important officials when the data surpassed a 95% confidence interval threshold. Officials would then log onto the RODS interface to analyze the data and determine a course of action.
- A two-tier review system was put in place to minimize the effect of false alarms.

Privacy Issues

- Intrusive chart reviews okayed for 2002 Winter Olympics.
- RODS was a “research project” so it needed HIPAA/IRB.
- Data use agreements with health systems.
- By law much of the information had to be protected, limiting the effectiveness of the surveillance. Also, RODS’ 7 syndromes did not match those 11 approved by Utah. Even with its NLP base, this was another possible limitation to the RODS system.
- Federal law now puts great emphasis on HIPAA. Minimal (HL7) data can now be used for public health surveillance.
- Trusted broker model implemented between UDOH, the RODS lab, and the individual health care systems. For IHC, UDOH was cut out of this relationship!

RODS in UTAH Week Seven

ID	Task Name
1	GOAL DATE FOR NETWORK AND DATA FEEDS
2	GOAL DATE FOR APPLICATIONS
3	OPENING CEREMONY
4	RODS PRESIDENTIAL DEMONSTRATION
5	**PUBLIC HEALTH LEVEL (Gesteland, Rolfs)**
6	**UU**
7	UU MOU review and sign
8	Pitt MOU review and sign
9	Utah Health Dept. R&S
10	IRB UU approval
11	Milestone: UU PERMISSION
12	**IHC**
13	IHC MOU review and sign
14	Pitt MOU review and sign
15	IRB IHC and ISC approval
16	Milestone IHC Permission
17	*Create and convene governing board*
18	Integrate into public health practice
19	**NETWORK LEVEL (Matyas)**
20	Finalize Olympic network designs
21	Get bldg, floor, and room loc. for lines
22	Order eqpt (leased line and VPN eqpt)
23	Order leased line circuits
24	**VPN**
25	preconfiguration and ship
26	**UU**
27	pix arrives
28	install and test
29	Milestone: UU VPN solution
30	**IHC**
31	pix arrives
32	install and test
33	Milestone: IHC VPN solution
34	**Leased Lines**
35	circuits delivered by AT&T
36	**UU**
37	Install and test
38	Milestone: UU 'Olympic' network
39	**IHC**
40	Install and test
41	Milestone: IHC 'Olympic' network

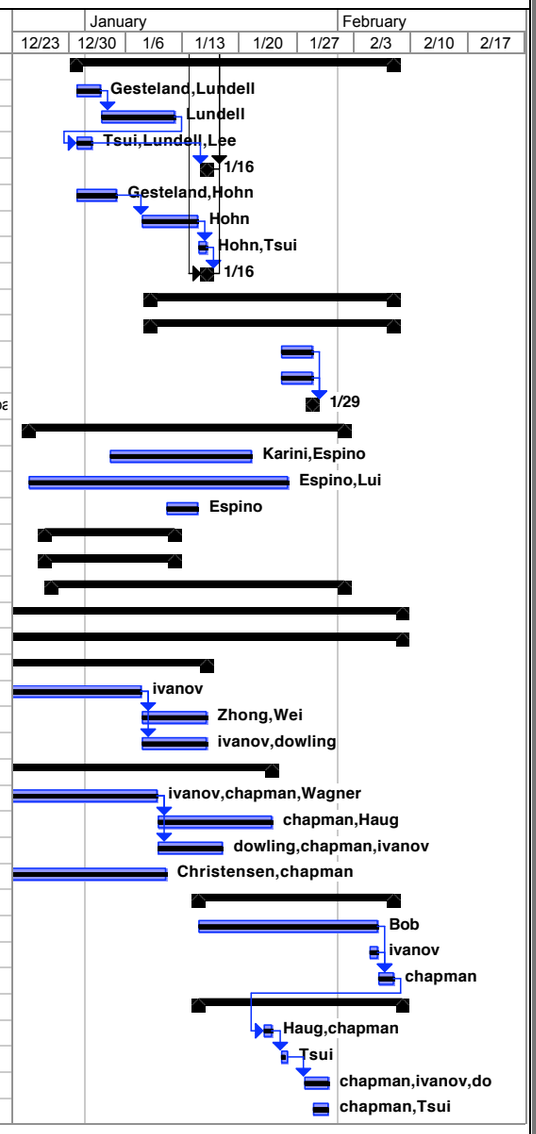


Project: UTAH RODS project
Date: Mon 2/11/02

Task
Split
Progress
Milestone
Summary
Project Summary
External Tasks
External Milestone
Deadline

RODS in UTAH Week Seven

ID	Task Name
42	**DATA LEVEL (Tsui)**
43	UU confirm data avail in HL7
44	UU any data enrichment
45	UU test messages
46	Milestone: UU data feeds
47	IHC confirm data avail in HL7
48	UU any data enrichment
49	IHC test messages
50	Milestone: IHC data feeds
51	**Add MRN to HL7 Messages for lookback**
52	**? MR Number encryption**
58	UU add MRN to HL7 msg
59	IHC add MRN to HL7 msg
60	Milestone: Data and UU IHC work done for lookba
61	**APPLICATION LEVEL (Wagner)**
62	GIS adaptation
63	User interface adaptation
64	User accounts
65	**Detection algorithm adaptation**
66	**Get historical data**
69	**Rapid Investigation capability**
78	**NLP CC-option (Ivanov)**
79	**development**
80	**Training preparation**
81	Training files - Map CC to syndrome
82	Automatic translation
83	Re-map CCs mapping to >1 syndrome
84	**Network preparation**
85	Design and create Netica network
86	Modify for new trainng file
87	Synonyms
88	Training tool
89	**Testing of Network**
90	10 fold cross-validation
91	ROC curves for network testing
92	Graph and Summarize results
93	**Implementation**
94	decide whether to go ahead
95	Install Mac
96	Install training tool and train newtork
97	Install trained version of INLS



Project: UTAH RODS project
Date: Mon 2/11/02

Task
Split
Progress
Milestone
Summary
Project Summary
External Tasks
External Milestone
Deadline

Results

Table 1 ■ Utah RODS Acute Care Visit Counts by Syndrome and Facility Type for the Period of February 8, 2002, through March 31, 2002

Syndrome	Urgent Care (%)	Emergency (%)	All Acute Care (%)
Constitutional	19,245 (27.23)	6,078 (13.85)	25,323 (22.10)
Respiratory	4,440 (6.28)	4,967 (11.32)	9,407 (8.21)
Encephalitic	2,895 (4.10)	4,107 (9.36)	7,002 (6.11)
Gastrointestinal	906 (1.28)	2,173 (4.95)	3,079 (2.69)
Rash	1,535 (2.17)	561 (1.28)	2,096 (1.83)
Hemorrhagic	285 (0.40)	762 (1.74)	1,047 (0.91)
Botulinic	18 (0.03)	30 (0.07)	48 (0.04)
None	41,360 (58.51)	25,217 (57.45)	66,577 (58.11)
Total	70,684	43,895	114,579

Results

- Over 51-day period, 233 logins from 31 unique users.
- Viewed Main screen 1244 times, 702 Epiplots, 511 Mapplots.
- Detection algorithm “fired” twice. Second instance interesting: 33 hemorrhagic complaints over 7 counties in 24 hours.

Discussion

- Olympics was definitely a special situation that drove the deployment of the RODS system. Without such a driving force, deployment could take years.
- Institutional data sharing a political not technical hurdle. With RODS system already largely developed, the implementation of the Trusted Broker consumed the most time and effort.
- Many natural hurdles (cost, experience, interoperability, data complexity, etc) were addressed or even eliminated prior to implementation, thus making the RODS deployment more easily successful.

Questions