International Society for Disease Surveillance Conference 2010

Track 4: Informatics

3. Automatically tracking diabetes using information in physicians' notes
RS Bhatia, S McClinton, and RF Davies
5. The evaluation of influenza surveillance data elements for the health information exchange minimum data set
H-G Chang, J-H Chen, S Jin, and PF Smith
7. Challenges in adapting a natural language processing system for real-time surveillance
WW Chapman, M Conway, JN Dowling, F-C Tsui, Q Li, LM Christensen, H Harkema, T Sriburadej, and JU Espino
9. Automated detection of data entry errors in a real time surveillance system
L Chen, A Dubrawski, N Waidyanatha, and C Weerasinghe
11. PHIN-MS deployment acceptability survey for the data transfer of syndromic data between hospitals and public health
YE Cheng, H Dhotre, and D Drociuk
13. Collaborative development of use cases for geospatially enabling a health information exchange
KF Comer, S Wiehe, B Dixon, R Gamache, J Wilson, and S Grannis
15. Developing an application ontology for mining clinical reports: the extended syndromic surveillance ontology
M Conway, J Dowling, R Tsui, and W Chapman
17. Real-time surveillance of influenza/pneumonia deaths: new strategies using grid computing and natural language processing
K Davis, C Staes, R Price, J Duncan, S Igo, and JC Facelli
19. Clinical decision support at the time of an e-prescription can sustainably decrease unwarranted use of antibiotics for acute respiratory infections
GB Rattinger, CD Mullins, E Onukwugha, IH Zuckerman, LD Walker, B South, A Gundlapalli, M Samore, and S DeLisle
21. Standardization to aid interoperability between NLP systems
G Divita, QZ Treitler, SM Meystre, B South, S Shen, R Cornia, J Garvin, M Samore, J Nebeker, S DuVall, J Potter, S Lee, P Haug, B Bray, W Chapman, H Harkema, L D'Avolio, P Elkin, M Tuttle, G Savova, A Coden, M Tanenblatt, I Sominsky, C Clark, J Friedlin, D Demner-Fushman, TC Rindflesch, A Aronson, AC Browne, M Fiszman, and O Bodenreider
23. Integrating medical examiner data in Utah
Jeffrey Duncan and Todd Grey
25. Applying classification and anomaly detection techniques to real-world data
W Edwards, A Vaid, and I Brooks
27. EpiScape: a map generation service for spatial temporal visualization
JU Espino and K Bui
29. Securing protected health information in NC DETECT
DM Falls, M Li, and A Waller
31. Improved diagnosis of group A streptococcal pharyngitis using real-time biosurveillance
AM Fine, V Nizet, and KD Mandl
33. Population-based, pathogen-specific surveillance in Utah
P Gesteland, K Korgenski, K Peterson, and M Samore
35. Placing surveillance in a preparedness business process framework
PJ Gibson and F Theodore
37. EHR-based syndromic surveillance during the 2009–2010 H1N1 pandemic
KS Fiske, M Siefert, and M Hoffman
39. Virtual public health informatics training: development and delivery of an online public health informatics course
Y Jiang and CJ Staes
41. Assessing address data quality for public health surveillance in Montreal
E Lanktree, C Jauvin, L Bédard, and D Buckeridge

43. SAGES: a suite of freely available software tools for electronic disease surveillance in resource-limited settings
SL Lewis, BH Feighner, WA Loschen, RA Wojcik, JF Skora, JS Coberly, and DL Blazes
45. Health IT meaningful use: how the EHR certification may improve the electronic laboratory reporting (ELR)
N Lipskiy, N Garrett, and S Tandon
46. Reporting nationally notifiable conditions (NNC): vocabulary aspects
S Ganesan, N Lipskiy, R Merrick, and C Vinion
47. Improving quality and interoperability of public health messaging guides: CDC messaging quality framework (MQF)
N Lipskiy, G Kesarinath, M Pray, and V Nwadiogbu
48. Influenza messaging and interoperability: the PHLIP assistance team approach
J Lipsky, L Cohen, M Meigs, and P Zarcone
50. Real-time, reusable, dynamic public health surveillance
P Mirhaji, R Vedam, and MM Vagnoni
52. HAISS Data Warehouse (HDW)—A new data access architecture for ESSENCE in the VA
A Mostaghimi, G Oda, C Lucero, P Schirmer, J Lombardo, R Wojcik, and M Holodniy
54. Implementing an outbreak/event tracking monitor for public health
S Mottice, J Reid, T Jeppson, and R Rolfs
55. Classification of errors for quality assurance in the emerging infections program influenza hospitalizations surveillance system
A Pérez, T D'Mello, L Kamimoto, L Finelli, and MH Torres-Urquidy
56. BioSINE: an intuitive visualization tool to enhance collaboration between research and practice in disease surveillance
P Picciano, J Beaubien, G Mizrahi, J Owen, F Grant, and AM Hegedus
58. Mitigating data collection challenges with adaptive frameworks
A Poku, RA Wojcik, J Skora, S Lewis, and DL Blazes
60. The National Biosurveillance Integration System
H Park, E Pedersen, D Stetson, D Carr, D Bailey, M Stephens, J Best, A Marouf, and T Quitugua
61. Development of public health communication tools using open source methods
Jonathan Reid, Susan Mottice, Wu Xu, and Matthew Samore
63. Identification of features for detection and prediction of homelessness from VA clinical documents
S Shen, BR South, M Palmer, S Duvall, M Samore, and AV Gundlapalli
64. Using clinician mental models to guide annotation of medically unexplained symptoms and syndromes found in VA clinical documents
BR South, M Palmer, S Shen, G Divita, S Duvall, M Samore, and AV Gundlapalli
66. Document classification toward efficient event-based biosurveillance
M Torii, N Nelson, and D Hartley
68. Building an automated Bayesian case detection system
F-C Tsui, J Espino, T Sriburadej, H Su, and J Dowling
70. Epibasket: a prototype web tool aimed to speed up and standardize the epidemiological investigation of an outbreak
W Xing, G Hejblum, GM Leung, and A-J Valleron
72. Map application to the New York state electronic syndromic surveillance system
C Weng, J-H Chen, C DiDonato, and H-G Chang
74. Enhance public health data capacity through statewide health information exchange of immunization records
W Xu, F Lanier, and N McConnell
76. Public health data sharing policy and informatics initiatives at China CDC
Y Zhang and Q Chen

ABSTRACT

Automatically tracking diabetes using information in physicians' notes

RS Bhatia, S McClinton, and RF Davies

Cardiology Research, University of Ottawa Heart Institute, Ottawa, Ontario, Canada
E-mail: rbhatia@ottawaheart.ca

Objective

This paper describes DMReporter, a medical language processing system that automatically extracts information pertaining to diabetes (demography, numerical measurement values, medication list, and diagnoses) from the free text in physicians' notes and stores it in a structured format in a MYSQL database.

Introduction

Patient consultations recorded as voice dictations are frequently stored electronically as transcriptions in free text format. The information stored in free text is not computer tractable. Advances in artificial intelligence permit the conversion of free text into structured information that allows statistical analysis.

Methods

DMReporter uses machine learning and natural language processing tools to do information extraction. Numerical measurement values extracted include blood pressure,

weight, LDL, HDL, total cholesterol, HbA1C, fasting glucose, glucose (unspecified), and creatinine. These values are present as multiple readings, potential target values, values over a period of time, varied location, values reflecting family history, changes in value versus absolute value, and so on. The methods developed in Bhatia *et al.*¹ extract and disambiguate these values while producing informative label-value pairs. The system extracts 290 medications in six categories using regular expressions and edit distance algorithm. The diagnosis detection uses negation detection in the sentences that mention diabetes using the NegEx algorithm.² The dataset used by the program consists of

Diabetes Report Card University of Ottawa Heart Institute/The Ottawa Hospital			
Patient Info		LAB Data	
Patient ID	10000001	Blood Pressure	May 13, 2009 140/80
Diabetic	Yes	Creatinine	May 13, 2009 92
YOB	1963	Fasting Glucose	May 13, 2009 6.4
Gender	Female	HDL	Oct 08, 2008 1.06
Age (last Visit)	46	LDL	Oct 08, 2008 2.5
		Weight	May 13, 2009 155
Measure		Medication	
	Score		
ACE/ARB	1	Entrophon	May 13, 2009 81 mg
Blood Pressure	0	Metformin	May 13, 2009 500 mg
LDL	0.5	Pravachol	May 13, 2009 40 mg
HbA1C/F. Glucose	1	Vasotec	May 13, 2009 10 mg
BMI	0		
Total	2.5/5		

Figure 1 Diabetes report card.

Figure 2 (a) Blood pressure recorded over time for a patient. (b) Metoprolol doses recorded over time for a patient.

81 932 transcribed outpatient notes collected from 30 459 patients over 5 years.

Results

Currently, the system automatically extracts diabetes diagnoses, medications along with dosage and frequency information and nine numerical lab values. It produces a report card for every patient (Figure 1), assigns a score out of five representing treatment and health, and tracks each variable extracted over the entire recorded period of patient history (Figures 2a and b) in addition to population reports.

Conclusion

DMReporter allows monitoring of diabetic patients' laboratory values and medications over the entire documented history of the patient. The solutions developed for this project can be applied to documents from other institutions and to other diseases. It demonstrates the ability to

develop patient-chart abstractors within a specified restricted domain.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

- 1 Bhatia R, Graystone A, Davies RA, McClinton S, Morin J, Davies RF, *et al.* Extracting information for generating diabetes report card from free text in physicians notes. NAACL HLT 2010 Second Louhi Workshop, 2010, pp 8–14. Available at <http://aclweb.org/anthology-new/W/W10/W10-11.pdf>.
- 2 Chapman W, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. Evaluation of negation phrases in narrative clinical reports. *Proc AMIA Symp* 2001, pp 105–14. Available at <http://www.ncbi.nlm.nih.gov/sites/ppmc/articles/PMC2243578/>.

ABSTRACT

The evaluation of influenza surveillance data elements for the health information exchange minimum data set

H-G Chang, J-H Chen, S Jin, and PF Smith

New York State Department of Health, Albany, NY, USA
E-mail: hgc04@health.state.ny.us

Objectives

Evaluate the availability, timeliness, and accuracy of MDS data elements received from one RHIO for emergency department (ED), in-patient, and outpatient visits. Compare the characteristics of patients meeting the HIE influenza-like illness definition who were admitted to the hospital or expired versus those discharged home.

Introduction

The American Health Information Community Harmonized Use Case for the Biosurveillance minimum data set (MDS) was implemented to establish data exchange between regional health information organizations (RHIOs) and the New York State Department of Health (NYSDOH) for accelerating situational awareness through the Health Information Exchange (HIE) Project. However, the completeness, timeliness of the reporting and quality of data elements in the MDS through RHIOs are still unknown and need further validation before we can utilize them for NYSDOH public health surveillance.

Methods

In-patient, outpatient, and ED visit data from one hospital in RHIO A between June and December 2009 were extracted. MDS data elements were evaluated for their availability and timeliness. Patients who had discharge disposition coded as admitted to the hospital or expired were selected as cases, and a random sample of age and patient-class matched patients with disposition coded as discharge home were selected as controls. Using chart review as gold standard, MDS data elements were evaluated for accuracy and completeness. Characteristics of patients with influenza, influenza-like illness, and pneumonia were compared for those admitted or expired versus non-admitted.

Results

There were 10,808, 15,429, and 17,901 patients seen from in-patient, outpatient, and ED, respectively, during the study period. In all, 74 patients from ED visits were admitted to hospitals, 38 died while in the ED, and 15,547 patients were

discharged home. There were 11,623 outpatients discharged home, 58 were admitted to hospitals, and one patient died. In all, 301 of the 10,718 in-patients died while hospitalized. The availability of MDS data elements is listed in Table 1. The median days from patient visit date to NYSDOH received date is 1 day for ED, and 2 days for in-patients and outpatients; the mean number of days were 5, 7, and 6

Table 1 Percentage of availability and accuracy of MDS data elements

MDS data elements availability	In-patient (%)	Outpatient (%)	ED (%)
Age	100	100	100
Gender	100	100	100
Zip code	100	100	100
Symptoms onset date	0	0	0
Chief complaint: free text	100	94	100
Diagnosis: free text	100	90	95
Diagnosis: ICD9	97	76	95
Discharge disposition	94	77	96
Temperature	0	0	0
Pulse oximetry	0	0	0
<i>MDS accuracy</i>			
Age	100	100	100
Gender	100	100	100
Zipcode	96	96	87
Diagnosis:ICD9	34	91	93
Discharge disposition	76	94	94

Abbreviations: ED, emergency department; MDS, minimum data set.

Table 2 Influenza patient characteristics by admitted or expired versus discharge home

Characteristics	Admitted or expired N (%)	Discharge home N (%)	P-value
Mean age	64	54	0.16
Temperature $\geq 100^{\circ}$ F	7 (35)	13 (41)	0.69
Sore throat	2 (11)	4 (12)	0.86
Cough	8 (42)	15 (45)	0.81
Shortness of breath	12 (60)	11 (33)	0.05
Myalgia	0 (0)	6 (18)	0.05
Sepsis	3 (17)	0 (0)	0.04
Pneumonia/ARD	11 (58)	9 (27)	0.03

days for ED, outpatient, and in-patient, respectively. After excluding duplicate records, a total of 189 patients admitted to hospitals or expired were selected as cases; 195 patients matched by 5-year age groups and patient class were selected as controls. The accuracy of MDS data elements is listed in Table 1. There were 24 cases and 33 controls had influenza-like illness, 13 cases were admitted to ICU. Characteristics of cases and controls were shown in Table 2.

Conclusion

Most of the MDS data elements were available and complete except onset date, temperature, and pulse oximetry. HIE

data (50%) were received within 2 days of visit date. The discrepancy of zip code, diagnosis code, and discharge disposition were due to multiple messages in the HIE data. Influenza patients died or admitted were more likely to be admitted to ICU and to have complications.

Acknowledgements

This paper was presented as an oral presentation at the 2010, International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

ABSTRACT

Challenges in adapting an natural language processing system for real-time surveillance

WW Chapman, M Conway, JN Dowling, F-C Tsui, Q Li, LM Christensen, H Harkema, T Sriburadej, and JU Espino

Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
E-mail: wendy.w.chapman@gmail.com

Objective

Adapt an existing natural language processing (NLP) system to be a useful component in a system performing real-time surveillance.

Introduction

We are developing a Bayesian surveillance system for real-time surveillance and characterization of outbreaks that incorporates a variety of data elements, including free-text clinical reports. An existing NLP system called Topaz is being used to extract clinical data from the reports. Moving the NLP system from a research project to a real-time service has presented many challenges.

Methods

We describe the approaches we are taking to address the challenges, along with results of the current implementation on findings relevant for Shigellosis and influenza surveillance.

Modeling relevant knowledge

Which diseases should we monitor, what clinical information provides evidence for those diseases, how can we represent the knowledge so that collaborators in public health, clinical medicine, knowledge engineering, and software development have access to consistent information? We are developing an application for ontology/thesaurus-relating diseases of interest to public health practitioners to relevant findings potentially described in clinical reports.¹ The thesaurus provides a single point of reference used to manually design a Bayesian case detection model and represent target concepts the NLP system will extract.^{2,3}

Creating an efficient web service for the NLP system

We are receiving real-time HL-7 messages from the University of Pittsburgh Medical Center and must process reports as they become available; however, the NLP system was much too slow for real-time processing. We increased efficiency by only loading a small portion of the UMLS Metathesaurus and by replacing the MetaMap indexing

module with an implementation of IndexFinder, which is substantially faster. Moreover, we lacked a communication mechanism between the NLP system and the Bayesian case detector. We created an xml schema based on the concepts in the thesaurus for standardized NLP output that can be parsed by the case detector.

Building a framework for tracking performance and diagnosing errors

Assessing errors and updating the NLP system's lexicon and rules were initially accomplished through a painstaking process involving translating output to tab-delimited files, calculating outcome measures in Excel, identifying reports with errors from the Excel files, and examining sentences causing the errors in separate text documents. We are building a framework to simplify results review that automatically calculates agreement between two xml files (using the schema described previously). To quickly identify errors, a user can click on a cell in the contingency table and find the reports contributing to the cell's number. Clicking on a report brings up the text of the report and shows the annotations made by the human reference standard and by the NLP system.

Results

Using the results review analysis tool we have iteratively identified errors on 53 target concepts and made changes to the NLP system using a development set of 26 laboratory-verified Shigellosis cases and 20 laboratory-verified influenza cases. Positive predictive value (PPV) and sensitivity are 84 and 47% for Shigellosis concepts and 89 and 67% for influenza concepts. Sensitivity is consistently lower than PPV, indicating an incomplete lexicon. We are developing a lexicon-building tool to mine synonyms from clinical documents and the UMLS Metathesaurus.

Conclusions

Implementing an NLP system for a real-time surveillance application is more difficult than we expected it to be, and the

challenges are caused less by NLP technicalities than they are by the lack of a robust environment for NLP engineering, adaptation, and improvement. We are developing processes for knowledge engineering and modeling, communication standards for applications using NLP system output, and a dashboard-like interface for performing results review, error analysis, and results tracking. We hope the tools and standards will be generalizable to other surveillance projects and can be used for NLP system adaptation in other domain areas.

Acknowledgements

The project is funded by CDC grants P01 HK000086 and 1U38 HK000063-01. This paper was presented as an oral presentation at the 2010, International Society for Disease

Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

- 1 Conway M, Dowling J, Chapman W. *Developing a Biosurveillance Application Ontology for Influenza-Like-Illness*. Proceedings of the OntoLex Workshop: Beijing, 2010.
- 2 Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp* 2001; 17–21. Available at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2243666/>.
- 3 Zou Q, Chu WW, Morioka C, Leazer GH, Kangarloo H. IndexFinder: a method of extracting key concepts from clinical texts for indexing. *AMIA Annu Symp Proc* 2003; 763–7. Available at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1480259/>.

ABSTRACT

Automated detection of data entry errors in a real time surveillance system

L Chen¹, A Dubrawski¹, N Waidyanatha², and C Weerasinghe²

¹Auton Lab, Carnegie Mellon University, Pittsburgh, PA, USA; and ²LIRNEAsia, Colombo, Sri Lanka
E-mail: lujiec@andrew.cmu.edu

Objective

We present a method for automated detection of systematic data entry errors in real time biosurveillance.

Introduction

Real-Time Biosurveillance Program (RTBP) introduces modern surveillance technology to health departments in Sri Lanka and Tamil Nadu, India.¹ Triage data from each patient visit (basic demographics, signs, symptoms, preliminary diagnoses) is recorded on paper at health facilities. Case records are transmitted daily to a central database using the RTBP mobile phone application. It is done by medical professionals in India, but in Sri Lanka, due to staffing constraints, the same duty is performed by lower cost personnel with limited domain knowledge. That results in noticeable differences in data entry error rates between the two locations. Most of such issues are due to systematic and subjective misinterpretations of the handwritten doctor notes by the data entry personnel. If not identified and remedied quickly, these errors can adversely affect accuracy and timeliness of health events detection. There is a need to support system managers in their efforts to maintain high reliability of data used for public health surveillance.

Methods

To address the need, we develop algorithms for automated detection of systematic data quality issues. They are used in automated, on-line screening of incoming data for potential discrepancies. Lists of potential issues are presented to human operators for evaluation. The operators may then correct the contents of the database, re-execute analyses, which might have been affected by errors, and follow-up with data entry personnel to prevent similar issues from happening in the future.

One of such algorithms relies on the assumption that most disease outbreaks do not affect population served by just one health facility. We expect neighboring facilities to show at least some level of correlated activity. We use entropy as a measure of uniformity of the geographic distribution of disease. Spatially isolated outbreaks will be characterized with low values of entropy, whereas widespread events will show high entropies. The data used in our experiments contains 62,000 disease cases

from 13 locations in Sri Lanka covering 158 disease categories. Each data entry clerk is assigned to a specific health facility. Therefore, systematic subjective data entry errors tend to show as localized patterns ('spikes' or 'dips') with low spatial entropy. Our algorithm exhaustively searches data for such instances.

Results and conclusion

The Figure 1 below presents the results of analysis limited to notifiable diseases observed in Sri Lanka during a specific period of time. Horizontal axis denotes normalized entropy of disease geospatial distribution, vertical axis depicts relative frequency of diseases, and each point corresponds to one notifiable disease. Points in the upper left quadrant of the plot are likely the results of miscoding. These diseases show relatively high numbers of cases combined with a very limited spatial dispersion. Indeed, upon investigation, it was revealed that the spike of *Tetanus* resulted from mistakenly entering immunization records as if they were disease cases. The phantom outbreak of measles resulted from mistakenly recording lexically similar diagnoses of muscle pain. However, the unexpected spike of 150 cases of *fever greater than 7 days* observed at just one location in February and March of 2010, an ailment not typically seen in dry season, was a legitimate health event.

Figure 1 Entropy map of notifiable disease in Sri Lanka.

Low quality of data can easily invalidate analyses. Implementers of analytic systems must be prepared to face such challenges, especially in environments with limited resources. We have shown how to automatically detect certain types of avoidable systematic data entry errors in support of real time biosurveillance.

Acknowledgements

This work was supported in part by the International Development Research Centre of Canada (Award 105130)

and National Science Foundation under grant number 0911032. This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

Reference

- 1 Sabhnani M, Dubrawski A, Waidyanatha N. T-Cube Web Interface for Real-time Biosurveillance in Sri Lanka. Proceedings of the 2009 ISDS Annual Conference, Miami, Florida.

ABSTRACT

PHIN-MS deployment acceptability survey for the data transfer of syndromic data between hospitals and public health

YE Cheng, H Dhotre, and D Drociuk

Division of Acute Disease Epidemiology (DADE), SC Department of Health and Environmental Control (SC DHEC), Columbia, SC, USA
E-mail: chengye@dhec.sc.gov

Objective

The objective of this presentation is to review and evaluate the use of Public Health Information Network (PHIN)¹ Messaging Service (PHIN-MS)² for the data transfer of the syndromic data between hospitals and public health. Included is an overview of the methodology used for PHIN-MS, and a review of the usage, adoption, benefits, and challenges within the hospitals and public health agencies in South Carolina. A formal survey is planned with results discussed during the presentation of this manuscript.

Introduction

PHIN-MS can send and receive data securely and automatically. It is used by many hospitals in the state(s) to send data to the South Carolina Department of Health and Environmental Control (SC DHEC) for both our National Electronic Disease Surveillance System and our South Carolina Early Aberration Alerting Network syndromic surveillance system.

DHEC has also implemented the PHIN-MS Route-not-Read (RnR) hub to allow hospitals to easily poll/receive syndromic data reports back from SC DHEC. This enables a two-way communication between the health partners. A basic process flow diagram is shown in Figure 1 below.

Figure 1 PHIN-MS data transfer diagram.

DHEC began using the PHIN-MS in 2006 for several systems, including the syndromic system beginning in 2008. We initially used PHIN-MS version 2.6, but upgraded to version 2.8.01.

As of August 2010, DHEC was receiving daily syndromic data from 14 emergency departments. These hospitals are receiving data from DHEC using PHIN-MS RnR. Many hospitals are expected to enroll. Positive feedback is normally received, but a formal survey is in process.

Methods

DHEC provides all of the PHIN-MS installation files and technical support to the hospitals, with additional technical support provided by the Center for Disease Control and Prevention in an as needed basis. Most of the intensive IT deployment work is done at SC DHEC rather than at the hospital level, especially with the PHIN-MS RnR hub.

Multiple anecdotal comments regarding the ease of deploying PHIN-MS have been received by SC DHEC. This prompted the creation of a formal survey tool in order to systematically assess the PHIN-MS deployment process. This survey will be provided to hospitals participating in both our National Electronic Disease Surveillance System-based system and the South Carolina Early Aberration Alerting Network syndromic surveillance system.

The survey will focus on the following areas: hospital adoption, satisfaction, challenges, and benefits of using PHIN-MS.

Some questions included in the survey are:

- How satisfied is the hospital using PHIN-MS?
- How easy is the hospital to adopt PHIN-MS as the messaging tool for the data exchange?
- How easy is the PHIN-MS installation process for the hospital?
- How satisfied is the hospital with the technical support provided by SC DHEC?

- What are the top challenges the hospital has experienced during the implementation?

Results

A summary of the survey will be discussed. Barriers to recruiting hospitals into the system were mostly because of competing priorities within the healthcare setting. South Carolina does not mandate hospitals to report syndromic data. Once a hospital is recruited into the system, sending data through PHIN-MS at the hospital has proved successful so far.

Conclusions

PHIN-MS sends data securely, reliably, and automatically. The hospitals need only very little resources to implement and maintain, especially with the PHIN-MS RnR; health partners can exchange data in a two-way communication securely and easily. PHIN-MS has shown its potential to be the standard for data transfer of syndromic data, as well as

several other systems like ELRs, cancer registry, and so on. PHIN-MS will be a valuable tool for increasing the state and local hospital's surveillance systems implementation.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

- 1 The Centers for Disease Control and Prevention. Public Health Information Network. <http://www.cdc.gov/phn/>.
- 2 The Centers for Disease Control and Prevention. Public Health Information Network Messaging System. <http://www.cdc.gov/phn/phnms/>.
- 3 The Centers for Disease Control and Prevention. The Early Aberration Reporting System. <http://emergency.cdc.gov/surveillance/ears/>.

ABSTRACT

Collaborative development of use cases for geospatially enabling a health information exchange

KF Comer¹, S Wiehe², B Dixon³, R Gamache^{3,4}, J Wilson⁵, and S Grannis^{3,4}

¹The Polis Center, Indiana Center of Excellence in Public Health Informatics—IU School of Medicine, Indianapolis, IN, USA; ²Children's Health Services Research, Indiana Center of Excellence in Public Health Informatics—IU School of Medicine, Indianapolis, IN, USA; ³Regenstrief Institute, Indiana Center of Excellence in Public Health Informatics—IU School of Medicine, Indianapolis, IN, USA; ⁴IU School of Medicine, Indiana Center of Excellence in Public Health Informatics—IU School of Medicine, Indianapolis, IN, USA; and ⁵IUPUI Department of Geography, Indiana Center of Excellence in Public Health Informatics—IU School of Medicine, Indianapolis, IN, USA
E-mail: kfrederi@iupui.edu

Objective

This presentation describes a collaborative approach for realizing the public health potential of a geospatially enabled statewide health information exchange.

Introduction

Given the clear relationship between spatial contexts and health,¹ the Indiana Center of Excellence in Public Health Informatics (ICEPHI) aims to serve both the needs of public health researchers and practitioners by contextualizing the health information of large populations. Specifically, ICEPHI will integrate one of the nation's largest health information exchanges, the Indiana Network for Patient Care (INPC),² with well-established community information systems that collect, geocode, organize, and present integrated data on communities in Indiana and surrounding states, including data on public safety, welfare, education, economics, and demographics.³

Methods

To integrate more than 3 billion clinical results for more than 12 million patients in the INPC with rich, locally available contextual data, an ongoing process for geocoding all clinical data is being developed that links the addresses associated with electronic medical records to geographical coordinates and other useful geographical identifiers including census block group data. To leverage this new data, a collaborative team of information scientists and public health researchers, practitioners, and decision makers is developing use cases that reflect diverse public health needs. These use cases in turn are being used to refine system requirements. ICEPHI anticipated that requirements would vary based on the spatial context relevant to the public

health issues of interest^{1,4} and targeted sources of contextual information.⁵

Results

The use case requirements differ across different dimensions. These dimensions include accuracy of geocoding results, type of geographic identifier (for example, county, census tract, neighborhood), and date of geographical identifier (for example, 1990 versus 2000 census tract ID). Some use cases, such as targeting neighborhoods for clinical interventions, do not require the same level of geocoding precision as others, such as assessing whether proximity to environmental health hazards relates to an individual's risk of a particular health outcome. Street centerline addressing geocoding is sufficient for the first use case example, while property parcel geocoding is more desirable for the later example. The most commonly needed contextual data and associated requirements for allowing proper linkage were identified for initial prototype development. The prototype geocoding service will return geographical coordinates and block group ID based on the use of a composite geocoding method that uses multiple spatial reference layers. Some use cases will require that geocoding be restricted to the use of a more limited set of reference layers. The developed metadata protocol will allow the source of generated spatial attributes to be tracked and reported.

Conclusions

Enabling geospatial data within health information exchanges has great potential for supporting and advancing public health research and practice. Multi-sector collaboration on the development and evaluation of associated uses cases allows informed decisions on system integration,

allowing spatially aware research and practice to be more quickly realized.

The functionality of the ICEPHI geocoding service will be expanded as use cases are further developed and prioritized. Potential uses of de-identified and aggregated health information by community-based organizations will be considered in the future.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

- 1 Diez Roux AV, Mair C. Neighborhoods and health. *Ann N Y Acad Sci* 2010;**1186**:125–45.
- 2 Biondich PG, Grannis SJ. The Indiana network for patient care: an integrated clinical information system informed by over thirty years of experience. *J Public Health Manag Pract* 2004; (Suppl): S81–6.
- 3 Hayes RM. Informing our nation: Improving how to understand and assess the USA's position and progress. *Libr Info Sci Res* 2006;**28**:154–7.
- 4 Mobley LR, Kuo TM, Andrews L. How sensitive are multi-level regression findings to defined area of context?: a case study of mammography use in California. *Med Care Res Rev* 2008;**65**:315–37.
- 5 Schootman M, Sterling DA, Struthers J, Yan Y, Laboube T, Emo B, *et al.* Positional accuracy and geographic bias of four methods of Geocoding in Epidemiologic research. *Ann Epidemiol* 2007;**17**:464–70.

ABSTRACT

Developing an application ontology for mining clinical reports: the extended syndromic surveillance ontology

M Conway, J Dowling, R Tsui, and W Chapman

Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
E-mail: michaelambroseconway@gmail.com

Objective

To develop an application ontology—the extended syndromic surveillance ontology (ESSO)—to support text mining of ER and radiology reports for public health surveillance. The ontology encodes syndromes, diagnoses, symptoms, signs and radiology results relevant to syndromic surveillance (with a special focus on bioterrorism).

Introduction

Ontologies representing knowledge from the public health and surveillance domains currently exist. However, they focus on infectious diseases (infectious disease ontology), reportable diseases (PHSKb—retired) and internet surveillance from news text (BioCaster ontology), or are commercial products (OntoReason public health ontology). From the perspective of biosurveillance text mining, these ontologies do not adequately represent the kind of knowledge found in clinical reports. Our project aims to fill this gap by developing a stand-alone ontology for the public health/biosurveillance domain, which (1) provides a starting point for standard development, (2) is straightforward for public health professionals to use for text analysis, and (3) can be easily plugged into existing syndromic surveillance systems.

Methods

The extended syndromic surveillance ontology (ESSO) uses the syndromic surveillance ontology (SSO) as a starting point. SSO is an ontology of chief complaints and consensus syndrome definitions developed by representatives from 10 currently functional surveillance systems.¹ In developing the ESSO, we tripled the number of concepts represented and added important new relations (for example, *has_symptom*). All new clinical concepts and relations were identified by an infectious disease physician (JD).

The ontology consists of 300 clinical concepts, each mapped to one (or more) of eight syndromes (see Table 1). Each concept is associated with a concept *type* (for example,

Table 1 Syndromes represented with example concepts

Syndrome	Example concept
Rash	Measles
Hemorrhagic	Hematuria
Botulic	Slurred speech
Neurological	Seizures
Constitutional	Lethargy
Respiratory, sensitive	Sore throat
Respiratory, specific	Pneumonia
Gastrointestinal, sensitive	Dehydration
Gastrointestinal, specific	Vomiting
Influenza-like-illness	Prostration

Note that respiratory and gastrointestinal syndrome are subdivided into *specific* and *sensitive* syndromes.

Table 2 Types of clinical concepts represented (with examples)

Clinical concept type	Example concept
Diagnosis	Emphysema
Syndrome	Reactive airways
Symptom	Abdominal pain
Chest X ray finding	X ray pneumonia
Bioterrorism disease	Cholera

diagnosis, bioterrorism disease—see Table 2) and, where possible, mapped to terms from other, related knowledge sources (for example, BioCaster, UMLS). Additionally, we include textual patterns (in the form of regular expressions), which help map clinical concepts to their textual instantiations (and *vice versa*). We encoded this structure as an OWL (web ontology language) file using the Protégé-OWL ontology editor (the current standard for ontology development).

Results

ESSO consists of 300 clinical concepts organized according to their clinical category (for example, *diagnosis*, *symptom*, or

radiology finding). Further, these 300 concepts are linked together through a series of relations (for example, DISEASE has_symptom Symptom, SYMPTOM isRelatedTo SYMPTOM). Preliminary evaluation in the influenza-like-illness domain has shown that the ontology has good domain coverage.²

Conclusion

The ESSO is a light weight, easily embedded ontology application designed to provide reasoning support for syndromic surveillance of clinical reports using text mining.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Okhmatovskaia A, Chapman W, Collier N, Espino J, Buckeridge DL. SSO: the syndromic surveillance ontology. 2009 *ISDS Annual Conference abstract* (unpublished).
- 2 Conway M, Dowling J, Chapman W. Developing a biosurveillance application ontology for influenza-like-illness. Proceedings of the OntoLex workshop: Beijing 2010.

ABSTRACT

Real-time surveillance of influenza/pneumonia deaths: new strategies using grid computing and natural language processing

K Davis¹, C Staes¹, R Price², J Duncan³, S Igo^{1,2}, and JC Facelli^{1,2}

¹Department of Biomedical Informatics, Salt Lake City, UT, USA; ²Center for High Performance Computing, University of Utah, Salt Lake City, UT, USA; and ³Utah Department of Health, Salt Lake City, UT, USA
E-mail: kailah.davis@utah.edu

Objective

This poster presents the rationale for designing a real-time surveillance system, based on mortality data, using grid and natural language processing tools that will address the current barrier that coded death certificate data not being available for several months. To develop a public health tool that delivers a timely surveillance system for influenza and pneumonia, we integrated death certificates from the Utah Department of Health, analytical grid services, and natural language processing tools to monitor levels of mortality. This example demonstrates how local, state, and national authorities can automate their influenza and pneumonia surveillance system, and expand the number of reporting cities.

Introduction

Surveillance of deaths due to influenza and pneumonia using death records has the potential to be a relatively inexpensive¹ and quick approach to tracking and detecting influenza and respiratory illness outbreaks; however, presently such a system does not exist because of the time delays in processing death records: in Utah, as is similar elsewhere in the United States, coded death certificate data are typically not available for at least 1–3 months after the date of death, and coded national vital statistics data are not available until after 2–3 years.²

Advanced informatics methods and tools can be applied to address the problems described above. Grid architecture is a promising methodology for the public health domain because it may combine and analyze unrelated data existing from independent domains, and provide computational and analytic resources on demand.³ Natural language processing tool can be used to automate the coding of 'cause of death' free text and allows mapping to other standardized codes. Coded information about causes of death is needed to integrate death certificate information into real-time

surveillance tools in order to trigger alarms for public health response.

Our objective is to describe a new model for influenza surveillance using NLP tools and grid-enabled data and analytic services.

Methods and preliminary results

The Utah Department of Health has made available de-identified death records from 1989 to 2008. Analytic methods have been developed using R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria to mimic analysis performed by CDC to detect outbreaks. The available data were divided into fifteen years that were used as baseline data and five years that were used as to simulate a real-time data feed. The historic baseline data were used to establish baseline trends and computed threshold values, which were compared with counts to trigger automatic alarms with the simulated real-time feed.

Future work

A grid-based analytic service will be developed using the analytical tool described above. Also, we will explore the feasibility of creating a grid version of the National Library of Medicine's natural language processing tool, MetaMap(<http://mmtx.nlm.nih.gov/>), which can be used to access the simulated real-time feed of the death records and provide coded information from the 'cause of death' field. Data grid services may be developed to simulate death records located in different administrative domains. For example, access to death records in Utah and neighboring states to identify regional outbreaks.

Conclusions

We expect that the integration of electronic death records with grid-enabled analytic and NLP tools will result in timely and flexible applications for public health surveillance. This

new non-traditional public health surveillance system will allow aggregation across jurisdictions, dynamic monitoring of diseases based on regional or national threats, and allow resource-limited public health agencies to access and share advanced informatics tools.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Dubrow R, Sestito JP, Lalich NR, Burnett CA, Salg JA. Death certificate-based occupational mortality surveillance in the United States. *Am J Ind Med* 1987;**11**:329–42.
- 2 Center for Disease Control and Prevention. *Death Certificate (document on the Internet)*. Center for Disease Control and Prevention: Georgia, 2006. update 2006 Aug 26, cited 2010 Feb 09]. Available from: <http://www.cdc.gov/NCIPC/pub-res/nvdrs-coding/Death-Certificate.pdf>.
- 3 Staes CJ, Xu W, LeFevre SD, Price RC, Narus SP, Gundlapalli A, *et al.* A case for using grid architecture for state public health informatics: the Utah perspective. *BMC Med Inform Decis Mak* 2009;**9**:32.

ABSTRACT

Clinical decision support at the time of an e-prescription can sustainably decrease unwarranted use of antibiotics for acute respiratory infections

GB Rattinger^{1,2}, CD Mullins², E Onukwughu², IH Zuckerman², LD Walker², B South^{3,4}, A Gundlapalli^{3,4}, M Samore^{3,4}, and S DeLisle^{1,2}

¹Veterans Administration Maryland Health Care System, Baltimore, MD, USA; ²Departments of Medicine and Pharmaceutical Health Services Research, University of Maryland, Baltimore, MD, USA; ³VA Salt Lake City Health Care System, Salt Lake City, UT, USA; and ⁴Department of Medicine, University of Utah, Salt Lake City, UT, USA
E-mail: sdelisle@umaryland.edu

Introduction

Microorganisms resistant to antibiotics (ABX) increase the mortality, morbidity and costs of infections. In the absence of a drug development pipeline that can keep pace with the emerging resistance mechanisms, these organisms are expected to threaten public health for years to come. Because exposure to ABX promotes the development of bacterial resistance, health care providers have long been urged to avoid using antibiotics to treat conditions that they are unlikely to improve, including many uncomplicated acute respiratory infections (ARI). We asked if interposing clinical decision support (CDS) software at the time of electronic order entry could adjust ABX utilization toward consensus guidelines for these conditions.¹

Methods

The CDS was programmed in-house and introduced on January 2003 at the Maryland VA site (Intervention) but not at the Utah VA site (control). The CDS targeted two ABX, gatifloxacin and azithromycin (targeted ABX). Access to all other ABX was unrestricted (other ABX). To derive an enriched, population-based sample, we applied a previously validated² ARI surveillance algorithm ((one of 197 'respiratory' ICD-9 diagnostic code or new cough remedies or temp ≥ 38 °C) and (text analysis of the clinical note for non-negated ARI symptoms)) to EMR entries surrounding all outpatient visits ($n = 4.1$ million) during our study period (January 2002–December 2006). Flagged records ($n = 7000$) were then manually abstracted on the day of each index visit for all information required to establish guideline-defined ARI diagnoses and ABX treatment.

Results

A total of 3831 unique patients satisfied the consensus case definitions for pneumonias (537), bronchitis (2931), sinusitis

(717) and non-specific acute respiratory infections (145). For the two targeted ABX, the proportion of unwarranted prescriptions for ARI decreased from 22 to 3%, pre- (2002) vs post-intervention period (2003–2006, $P < 0.0001$). This proportion did not change for the other ABX at the intervention site (30 pre vs 31% post) or for both the targeted (16 vs 20%) and the other ABX (22 vs 27%) at the control site

(Figure, *upper panel*). The CDS did not impede utilization of antibiotics when they were indicated (*middle panel*), including all cases with pneumonia. Azithromycin and gatifloxacin remained the predominant antibiotics prescribed when indicated for ARI (68–70% of all ABX at both sites, pre and post, *lower panel*).

Conclusion

CDS interposed during the e-prescription process nearly extinguished unwarranted use of targeted ABX for the treatment of ARI for 4 years, and did so without reducing indicated use or shunting misutilization toward unrestricted agents. If deployed at the appropriate scale, this approach could exert a positive impact on public health.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

- 1 Gonzales R, Bartlett JG, Besser RE, Cooper RJ, Hickner JM, Hoffman JR, *et al.* Principles of appropriate antibiotic use for treatment of acute respiratory tract infections in adults: background, specific aims, and methods. (2001). *Ann Intern Med* 2001;134:479–86.
- 2 DeLisle S, South BR, Anthony J, Kalp E, Gundlapalli A, Curriero FC, *et al.* Combining free text and structured electronic medical record entries to detect acute respiratory infection. *PLoS ONE* 2010;5(10): e13377.

ABSTRACT

Standardization to aid interoperability between NLP systems

G Divita^{1,2}, QZ Treitler^{1,2}, SM Meystre^{1,2}, B South^{1,2}, S Shen^{1,2}, R Cornia^{1,2}, J Garvin¹, M Samore^{1,2}, J Nebeker^{1,2}, S DuVall^{1,2}, J Potter^{1,2}, S Lee², P Haug², B Bray², W Chapman³, H Harkema⁴, L D'Avolio⁵, P Elkin⁶, M Tuttle⁶, G Savova⁷, A Coden⁸, M Tanenblatt⁸, I Sominsky⁸, C Clark⁹, J Friedlin¹⁰, D Demner-Fushman¹¹, TC Rindflesch¹¹, A Aronson¹¹, AC Browne¹¹, M Fisman¹¹, and O Bodenreider¹¹

¹VA Salt Lake City Health Care System, Salt Lake City, UT, USA; ²University of Utah, Salt Lake City, UT, USA; ³University of California San Diego, San Diego, CA, USA; ⁴University of Pittsburgh, Pittsburgh, PA, USA; ⁵VA Boston Healthcare System, Boston, MA, USA; ⁶Mount Sinai, New York, NY, USA; ⁷Harvard Medical School, Boston, MA, USA; ⁸IBM T.J. Watson Research Center, Hawthorne, NY, USA; ⁹Mitre Corporation, Boston, MA, USA; ¹⁰Regenstrief Institute, Indianapolis, IN, USA; and ¹¹NLM, Bethesda, MD, USA
E-mail: guy.divita@hsc.utah.edu

Objective

The Consortium for Healthcare Informatics Research, a Department of Veterans Affairs (VA) Office of Research and Development is sponsoring the development of a standard ontology and information model for Natural Language Processing interoperability within the biomedical domain.

Introduction

There are a number of Natural Language Processing (NLP) annotation and Information Extraction (IE) systems and platforms that have been successfully used within the medical domain.¹ Although these groups share components of their systems, there has not been a successful effort² in the medical domain to codify and standardize either the syntax or semantics between systems to allow for interoperability between annotation tools, NLP tools, IE tools, corpus evaluation tools and encoded clinical documents. There are two components to a successful interoperability standard: an information and a semantic model.

Methods

Platform-specific information models, such as the UIMA CAS, GATE annotation graphs, and to a lesser degree, Protégé frames and Knowtator,¹ have been adapted to serve as the information model. However, each includes complexity and/or verbosity that has hindered wide adoption. A GATE-lite syntax is under development through this effort, with an *Annotation* at the center of the syntax.

Design principles have emerged around the representation of each *Annotation*: decouple the message syntax from the semantics; design for (space) efficiency; use a standoff annotation model. Create a model is simple enough to be

adopted by a wide community, yet be expressive enough to encode a clinical document, a named entity, relationships between entities, and be able to represent temporal features. This model should be usable within annotation tasks, information extraction tasks, and document and corpus evaluation tasks.

Each *Annotation* is intrinsically typed with a tag name or category from a tag set. The naming convention for tags and their semantics are the focus of the other component of the interoperability standard. Semantics have been a stumbling block of other efforts. Under specification of entities such as *Sentence*, *Phrase* and *Token* has led to incomplete integrations at best. This effort includes definition and naming conventions from several established standards to define document structure and named entity semantics. Clinical Document Architecture³ standard is being used to define clinical document structure, and to represent named entities at the annotation level. This is being augmented with tags from the Penn Treebank parsing and tagging guidelines⁴ to define types of phrases, and part-of-speech tags. The model that has emerged includes components such as *Token* and *Term* that have otherwise not been defined within adoptable standards. Elements of these standards have been augmented to allow for relationships between named entities, and for post-coordinated concept coverage within a named entity.

Discussion

A draft should be available for discussion in September 2010. A reference version 0.0 should be available shortly thereafter. This effort is not just an academic exercise for Consortium for Healthcare Informatics Research. The Veteran's Informatics and Computing Infrastructure initiative within

Veterans Affairs has a real need to have an interlingua between external NLP, annotations and IE systems, as well as for in-house development.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

- 1 Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J of Biomed Inform. Biomedical Natural Language Processing* 2009;**42**:760–72.
- 2 Ide NM, Suderman K. Bridging the Gaps: Interoperability for GrAF, GATE, and UIMA. *Proceedings of the Third Linguistic Annotation Workshop, ACL-IJCNLP 2009*, pp 27–34. www.aclweb.org/anthology/W/W09/W09-3004.pdf.
- 3 Dolin R, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, *et al.* HL7 clinical document architecture, release 2. *J Am Med Inform Assoc* 2006;**13**:30–9.
- 4 Penn Treebank Guidelines. <http://www.cis.upen.edu/~treebank>.

ABSTRACT

Integrating medical examiner data in Utah

Jeffrey Duncan, and Todd Grey

Utah Department of Health, Salt Lake City, UT, USA
E-mail: jduncan@utah.gov

Objectives

The objectives of the Utah Medical Examiner Database (UMED) project are:

- To provide a single point of entry for medical examiner pathologists and staff to manage investigation information.
- To develop an operational system that links death certificate, medical examiner, and laboratory data in real time as a resource for epidemiology and public health surveillance.

Introduction

The Office of the Medical Examiner (OME) is a statewide system for investigation of sudden and unexpected death in Utah. OME, in the Utah Department of Health (UDOH), certified over 2000 of the 13,920 deaths in Utah in 2008.

Information from OME death investigations is currently stored in three separate UDOH data silos that have limited interoperability. These three electronic data systems include death certificates, medical examiner investigations, and laboratory results. Without interoperability, OME staff is required to enter the same data into multiple systems. In addition, the process of requesting laboratory analysis and receiving results is paper based, significantly slowing final cause of death determination in a majority of cases.

Epidemiological studies and surveillance activities are hindered by the lack of systems integration in UDOH and often require retrospective data linkage. As an example, in 2005, CDC and the UDOH reported that deaths in Utah caused by drug poisoning from non-illicit drugs had increased fivefold from 1991 to 2003.¹ This significant finding relied on retrospective linkage of death certificates, medical examiner records, and toxicology results to describe the problem.

In 2008, funding from a bioterrorism grant from the US Department of Homeland Security was secured to support development of a unique, integrated system for medical examiner and death certificate data.

Methods

The UMED was designed by staff from Utah's Office of Vital Records and Statistics, OME, and Department of Technology Services. UMED was designed with a three-tier architecture. In the data tier, information from historical medical examiner investigations has been linked with death certificate data to form a foundation database with combined death certificate and investigation information from 1991 to present. Moving forward, deaths entered into UMED will be intrinsically linked to death records in the data layer. Laboratory data is linked in the logic tier where UMED's unique person identifier is used to send and receive requests from the laboratory information system.

For deaths not under OME jurisdiction, funeral directors and non-OME physicians will interact through the web-based electronic death registration system interface to file death certificates. OME staff and physicians will interact through the web-based UMED interface that will provide a single point of entry for investigation, death certificate and laboratory data, eliminating multiple entry.

Figure 1 Architectural diagram of the UMED system.

The UMED system is currently undergoing internal testing and is scheduled for user acceptance testing in September 2010 (Figure 1).

Conclusions

The UMED system represents a major step forward for epidemiology and surveillance in Utah. The UMED database, containing linked medical examiner and death certificate data from 1991 to the present, is a rich resource for retrospective epidemiological research. Intrinsically linked operational data from UMED will facilitate real-time surveillance by public health programs in the UDOH.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

Reference

- 1 Caravati EM, Grey T, Nangle B, Rolfs RT, Peterson-Porucznik CA. Increase in poisoning deaths caused by non-illicit drugs—Utah, 1991–2003. *MMWR* 2005;**54**:33–6.

ABSTRACT

Applying classification and anomaly detection techniques to real-world data

W Edwards¹, A Vaid², and I Brooks¹

¹National Center for Supercomputing Applications, Urbana, IL, USA; and ²Champaign-Urbana Public Health District, Champaign, IL, USA
E-mail: wedwards@ncsa.uiuc.edu

Objective

This paper compares different approaches with classification and anomaly detection of data from an emergency department.

Introduction

Real-world public health data often provide numerous challenges. There may be a limited amount of background data, data dropouts, noise, and human error. The data from an emergency department (ED) in Urbana, IL includes a diagnosis field with multiple terms and notes separated by semicolons. There are over 7000 distinct terms, excluding the notes. Because it begins in April 2009, there is not yet adequate background data to use some of the regression-based alerting algorithms. Values for some days are missing, so we also needed an algorithm that would tolerate data dropouts.

INDICATOR¹ is a workflow-based biosurveillance system developed at the National Center for Supercomputing Applications (NCSA). One of the fundamental concepts of INDICATOR is that the burden of cleaning and processing incoming data should be on the software, rather than on the health care providers.

Methods

There were two major challenges to processing the ED data. First, we needed a way to reduce the vocabulary to a more manageable size. Second, we needed an algorithm that could tolerate a limited amount of baseline data and some data dropouts.

We grouped the terms into six syndromic groups: 'GI-Sensitive', 'GI-Specific', 'Respiratory-Sensitive', 'Respiratory-Specific', 'Flu-Like Illness', and 'Constitutional'.² The 'sensitive' groups include a larger set of symptoms than the 'specific' groups. Generated graphs suggested elevated activity in the Respiratory and Flu-like Illness groups around the time of the H1N1 flu outbreak in Fall 2009.

To generate alerts, we used a modified version of CDC's Early Aberration and Reporting System (EARS).³ EARS uses an

Estimated Weight Moving Algorithm (EMWA) to generate alerts. The modified approaches expand the baseline period to 28 days, separate data into weekdays and weekends, and also adjust for the total number of ED visits on a particular day.

For comparison, we grouped the data into Flu-like Illness, Respiratory-Sensitive, Respiratory-Specific, GI-Sensitive, and GI-Specific, and ran the EARS algorithm on the raw data, segregated into weekends and weekdays, and adjusted for total number of visits.

Results

For the Flu-Like Illness Syndrome, the algorithm that separated weekend from weekday data and adjusted for the total number of ED visits was the most sensitive, generating alerts on 19 days (~6% of the total). This pattern also held for the respiratory-sensitive and respiratory-specific groups (Figure 1).

Conclusions

This approach has yielded promising results, and in the future, we plan to expand the number of syndromic groups to explore rates of ED activity related to substance abuse and

Figure 1 Number of alerts generated for each syndrome group using different modifications of the EARS algorithm. The blue color analyzes only the daily counts for a particular syndrome. The red color separates the counts into weekdays and weekends. The green color adjusts for the total number of ED visits that day. The purple color both separates weekday/weekend data and adjusts for ED totals.

West Nile virus. The modified EARS algorithms also worked well for us, and we plan to apply them to the school absence data, where enrollment, as well as absence figures have fluctuated.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Brooks I, Edwards W. INDICATOR: A Cyberenvironment for biosurveillance and response. *Syndromic* 2009, 8th Annual Conference of the International Society for Disease Surveillance, 2009.
- 2 Chapman W. Developing Syndrome Definitions based on Consensus and Current Use. *J Am Med Inform Assoc* 2010;17: 595–601.
- 3 Tokars J, Burkom H, Xing J, English R, Bloom S, Cox K, *et al.* Enhancing Time-Series Detection Algorithms for Automated Biosurveillance. *Emerging Infectious Diseases* 2009; 15:4.

ABSTRACT

EpiScape: a map generation service for spatial temporal visualization

JU Espino, and K Bui

Real-time Outbreak and Disease Surveillance Laboratory, University of Pittsburgh, Pittsburgh, PA, USA
E-mail: jest4@pitt.edu

Objective

This paper describes EpiScape, our map generation service. It generates three-dimensional static or animated maps as Keyhole Markup Language (KML) files that can be used to display epidemiologic data over time and space using Google Earth or Google Maps software.

Introduction

The KML format has become a recognized standard for the distribution of geographic information system data.¹ In most recent versions of the Real-Time and Outbreak Disease Surveillance (RODS) system, we standardized on KML as our mapping solution. This decision obviates the need for commercial GIS servers and clients, and permits users to easily overlay RODS map output with other websites and

software that output KML, for example, EPA, NASA, and NOAA.

We quickly recognized that the mapping tools in RODS have broad applicability in public health and other domains where there is a requirement to display spatial temporal data as it relates to state, county, and zip code geographies. To facilitate these needs, we created the EpiScape map generation service for public use.

Methods

EpiScape comprises a spatial database, map generation server, and Google Earth. We utilize the open source PostgreSQL database to store the spatial data. The map generation server is implemented in Java Enterprise Edition and makes significant use of the PostGIS Java libraries. The

Figure 1 Stores monitored by the National Retail Data Monitor for over the counter medication sales aggregated by county as of 22 April 1999.

client for viewing the output of the EpiScape service is Google Earth. Both free and commercial versions of Google Earth are compatible with the EpiScape output.

We programmed the map generation server so that users can define the:

- title for their map
- spatial or spatial temporal data
- labels for each geographic area
- the number for bins and binning method
- a color scheme
- a map type
- transparency levels
- the magnitude of the three-dimensional effect
- the degree of detail for polygons

The service is accessible from a web page or through an http-based application programming interface (API) (API documentation is available at http://betaweb.rods.pitt.edu/wiki/index.php/KML_Generator). For confidentiality, the service deletes map data immediately after the users download their file. In addition, the source code for the service is available from the RODS Open Source Project.

We deployed the EpiScape service on the Apache Tomcat Servlet Engine and Apache web server. The service utilizes SSL encryption when transferring data to and from the user. We loaded the PostgreSQL database with state, county, and

zip code tabulation area polygon data from the United States Census.

Results

We use the EpiScape service as the map generation tool for multiple software projects at our laboratory. These projects include the RODS system, National Retail Data Monitor, and Allegheny County influenza monitoring system. Figure 1 is an example map from the EpiScape service showing the number of stores monitored by the National Retail Data Monitor by county.

Conclusions

The EpiScape map generation service has become an indispensable tool at our laboratory. Its web accessibility makes it easy for users to create high quality maps with minimal effort. As it has a web-based API, we have been able to easily incorporate it into other software projects. We hope that users outside our laboratory will find it useful.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

Reference

- 1 <http://www.opengeospatial.org/standards/kml/>.

ABSTRACT

Securing protected health information in NC DETECT

DM Falls, M Li, and A Waller

Carolina Center for Health Informatics, Department of Emergency Medicine, UNC, Chapel Hill, NC, USA
E-mail: dfalls@med.unc.edu

Objective

This paper describes how the North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) utilizes various methods of encryption and access control to protect sensitive patient data during both integration and reporting.

Introduction

NC DETECT receives daily data files from emergency departments (ED), the statewide EMS data collection system, the statewide poison center, and veterinary laboratory test results. Included in these data are elements, which may contain Protected Health Information (PHI). It is the responsibility of NC DETECT to ensure that security of these data is managed during their entire life cycle, including receiving, loading, cleaning, storage, managing, reporting, user access, archiving, and destruction. A web interface is provided for users at state, regional and local levels to access syndromic surveillance reports, as well as reports for broader public health surveillance such as injury, occupational health, and disaster management.

Methods

Data files are downloaded via the internet using both Secure File Transfer Protocol (SFTP) and Secure Hyper Text Transfer Protocol (HTTPS). An off-the-shelf Extraction, Transformation and Loading (ETL) tool, capable of receiving data from any of nearly 200 data types including ASCII, HL7 and XML, allows for easy database loading and data encryption. A combination of secure hash algorithm (SHA-1) and triple DES encryption algorithm are used to secure PHI upon database loading. Role based access with multiple tiers controls, data source, geography, aggregate data, line listing data, PHI, and annotation privileges, functionality which allows authorized users to document signals and keep track of signal investigations.

	Hospital Emergency Department	Poison Center	PreMIS
Medical Record Number	Hash & Encrypt	N/A	N/A
Account Number	Hash & Encrypt	N/A	N/A
Notes	N/A	Encrypt	N/A
Last Name	N/A	N/A	Encrypt
First Name	N/A	N/A	Encrypt
Middle Name	N/A	N/A	Encrypt
Incident Address	N/A	N/A	Encrypt
Patient Address	N/A	N/A	Encrypt

Figure 1 Security methods per data source/data element.

Data files are retained for 14 days in a location secured using Windows Encrypted File System (EFS).

Results

NC DETECT currently receives data from 120 hospital-based emergency departments, 100 EMS systems, and the state poison control center. Eight various data elements are either encrypted, hashed or both (Figure 1).

There are 356 active web interface users; 276 are authorized to access limited PHI based on professional role and geographical location.

Conclusion

PHI must be secured for both storage and transportation. NC DETECT's data processing system provides the functionality to meet HIPAA standards for data storage encryption¹ and our Role Based Web Interface provides protection of data being transmitted across the internet.² NC DETECT 4.0 provides users at all levels with secure and tailored access to syndromic, injury, post-disaster, occupational health and other types of public health surveillance reports. Role-based

access designs must be flexible enough to accommodate changing user needs as well as state and federal privacy and security regulations.

Acknowledgements

Funding for NC DETECT is provided by the North Carolina Division of Public Health. This paper was presented as a

poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA on 1–2 December 2010.

References

- 1 45 CFR Part 142, § 142.308 (c). 'Technical security services to guard data integrity, confidentiality and availability.'
- 2 45 CFR Part 142, § 142.308 (d). 'Technical security mechanisms.'

ABSTRACT

Improved diagnosis of group A streptococcal pharyngitis using real-time biosurveillance

AM Fine^{1,2}, V Nizet^{1,2}, and KD Mandl^{1,2}

¹Informatics and Emergency Medicine, Children's Hospital Boston, Boston, MA, USA; and ²Department of Pediatrics, University of California San Diego, San Diego, CA, USA
E-mail: Andrew.Fine@childrens.harvard.edu

Objective

The objective of this study was to measure the value of integrating real-time contemporaneous local disease incidence (biosurveillance) data with a clinical score, to more accurately identify patients with Group A Streptococcal (GAS) pharyngitis.

Introduction

Group A Streptococcal (GAS) pharyngitis, the most common bacterial cause of acute pharyngitis, causes more than half a billion cases annually worldwide. Treatment with antibiotics provides symptomatic benefit and reduces complications, missed work days and transmission. Physical examination alone is an unreliable way to distinguish GAS from other causes of pharyngitis, so the 4-point Centor score, based on history and physical, is used to classify GAS risk. Still, patients with pharyngitis are often misclassified, leading to inappropriate antibiotic treatment of those with viral disease and to under-treatment of those with *bone fide* GAS. One key problem, even when clinical guidelines are followed, is that diagnostic accuracy for GAS pharyngitis is affected by earlier probability of disease, which in turn is related to exposure. Point-of-care clinicians rarely have access to valuable biosurveillance-derived contextualizing information when making clinical management decisions.

Methods

We analyzed data from patients tested for GAS, who presented with pharyngitis from 2006 to 2008 to Minute-Clinic, a large national retail health chain. Analysis was restricted to nine markets with >7000 patient visits for pharyngitis, for a total of 138,910 patient encounters across six states. Anonymized extracted data included visit date, location, signs and symptoms included in the Centor score, and pharyngitis test results. To enable integration of contemporaneous, local GAS data with clinical data, we created a biosurveillance variable called the recent local proportion positive (RLPP), a moving window reflecting the

proportion of local patients testing positive in the previous week. Patients were grouped by Centor score (0–4) and further categorized by RLPP. We calculated the percent of patients who tested positive for GAS for all combinations of Centor score and RLPP. Using standard metrics (sensitivity, specificity, AUC), we compared the accuracy of the Centor score alone and RLPP alone with the accuracy of a biosurveillance-responsive score that integrated the Centor score and the RLPP to predict which patients tested positive for GAS. We examined the public health effects of subtracting 1 point from the Centor score when the RLPP was below certain thresholds, and adding 1 point when the RLPP exceeded thresholds.

Results

There was no distinct seasonal GAS pattern. For patients with Centor scores of 1–4, represented by the top 4 lines in the figure below, the percent of patients testing positive increases as the RLPP increases ($P < 0.0001$). When RLPP > 0.30, managing patients with Centor scores of 1 (where the American College of Physicians recommendation is neither

test nor treat) as if their scores were 2 would identify 114,850 previously missed patients who would test positive for GAS each year in the United States while misclassifying 33,161 patients who tested negative. When RLPP <0.20, approaching patients with Centor scores of 3 (where one guideline suggests empiric treatment) as if their scores were 2 would spare unnecessary antibiotics for 78,367 patients while missing 8,195 positives. The AUC is best for the biosurveillance-responsive model incorporating RLPP with the Centor score (0.72), followed by Centor score alone (0.70), and then by RLPP alone (0.57).

Conclusions

Incorporating live epidemiological data into clinical guidelines for GAS should be considered to reduce missed cases when the contemporaneous incidence is elevated, and to spare unnecessary antibiotics when the contemporaneous incidence is low.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

ABSTRACT

Population-based, pathogen-specific surveillance in Utah

P Gesteland^{1,2}, K Korgenski², K Peterson², and M Samore¹

¹University of Utah School of Medicine, Salt Lake City, UT, USA; and ²Intermountain Healthcare, Salt Lake City, UT, USA
E-mail: Per.Gesteland@hsc.utah.edu

Objective

The objective of this study is to describe a system—Germ Watch—that provides information about the regional activity of common communicable infectious diseases.

Introduction

Epidemiological information realized by modern disease-surveillance systems offers great potential for supporting clinical decision-making. Providing health practitioners with population-based, pathogen-specific information about regional communicable infectious disease epidemiology can engender enhanced knowledge about specific pathogens, which may, in turn, lead to improved clinical performance. To enhance the pathogen-specificity of Utah's surveillance system, which includes tracking syndromes¹ and notifiable diseases, we developed a system that tracks microbiologic testing in Utah's largest health care delivery system.

Methods

Setting

Intermountain Healthcare operates 21 hospitals and >100 outpatient clinics. The Intermountain Healthcare Data Warehouse is a seven terabyte (TB) database containing clinical (lab, radiology, meds, vitals) and administrative data from inpatient and outpatient settings.

Data source

Microbiological testing, including molecular diagnostic testing (DFA, PCR, EIA), and cultures ordered during routine clinical care and performed in one of Intermountain Healthcare's microbiology laboratories.

Pathogens currently tracked

Nine respiratory pathogens and seven enteric pathogens for which testing is available and routinely performed (Table 1).

Data warehousing

Lab results are uploaded daily into a dimensional (star) schema in the Intermountain Healthcare Data Warehouse to

Table 1 Pathogens tracked by Germ Watch and current count of encounters in the database (2002 through Sept. 2010)

Pathogen	Encounters
<i>Respiratory</i>	
Influenza A ^a	13,958
Influenza B ^a	2108
Respiratory Syncytial Virus (RSV)	16,072
Human Metapneumovirus (hMPV)	3025
Adenovirus	2517
Parainfluenza (types 1, 2 and 3)	4130
Rhinovirus	7031
Enterovirus	914
Bordetella pertussis ^a	712
<i>Enteric</i>	
Rotavirus	2898
E. Coli 0157H7 ^a	973
Giardia lamblia ^a	1503
Cryptosporidium ^a	1286
Salmonella ^a	1160
Shigella ^a	210
Campylobacter ^a	1563

^aReportable disease in Utah.

support *ad-hoc* multidimensional online analytical processing analysis of daily surveillance data. Additional normalized tables are utilized to enable fast web-query performance of the web applications. Data is available at patient, region and organism granularity.

Reporting

We use the IBM COGNOS (International Business Machines Corp. Armonk, New York, USA; <http://www-01.ibm.com/software/data/cognos/>) platform to automatically generate and disseminate dashboard-style summary reports, and NetCharts (Visual Mining, Inc., Rockville, MD, USA; <http://www.visualmining.com/>) to provide interactive, web-based graphing. GIS visualizations are made available using Google Maps (Google Inc. 2011. Map data—Europa Technologies, INEGI).

Figure 1 Time-series graph of Germ-Watch data showing yearly epidemics of RSV (blue), Influenza (dark red) and hMPV (orange).

Results

Presently, the Germ-Watch schema contains over 60,000 encounters associated with a lab-detected infection (Table 1). Time-series graphs of these data provide meaningful information about epidemic and endemic activity of various pathogens (Figure 1).² Summary reports, along with a bulleted text summary of the week's activity, are e-mailed weekly to >300 physicians and made available on the web. The system has proven sustainable and has been well-received by Utah providers who believe that it helps their clinical performance.³

Conclusions

Pathogen-specific information derived from routine diagnostic testing can provide an important signal to population

health, when data are available at a population level. Systems that generate and disseminate this information to health practitioners can, and will have an important role in supporting public health practice and clinical decision-making.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Gesteland PH, Gardner RM, Tsui FC, Espino JU, Rolfs RT, James BC, *et al.* Automated syndromic surveillance for the 2002 Winter Olympics. *JAMIA* 2003;10:547–54.
- 2 Gesteland PH, Samore MH, Pavia AT, Srivastava R, Korgenski K, Gerber K, *et al.* Informing the front line about common respiratory viral epidemics. *Proc AMIA Symp* 2007; 274–8.
- 3 Gesteland PH, Allison MA, Staes CJ, Samore MH, Rubin MA, Carter ME, *et al.* Clinician use and acceptance of population-based data about respiratory pathogens: implications for enhancing population-based clinical practice. *Proc AMIA Symp* 2008; 232–6.

ABSTRACT

Placing surveillance in a preparedness business process framework

PJ Gibson¹, and F Theodore²

¹Marion County Health Department, Ocala, FL, USA; and ²CMA Consulting, Kansas City, MO, USA
E-mail: jgibson@hhcorp.org

Objective

The Public Health Emergency Preparedness Business Process Interdependencies Framework was developed through a collaboration of state and local health departments, brought together to define public health's business processes related to preparedness. This session will explain the framework, the role of surveillance within it, and the methodologies used to develop it.

Introduction

Most public health workers could reach agreement on the general functions and operations regarding public health's traditional operations, such as food safety inspections, outbreak investigations, or immunization clinics. However, there is much less shared understanding about public health's emergency preparedness activities.

A consortium of state and local health departments was brought together to define public health's business processes related to preparedness. That consortium developed the Public Health Emergency Preparedness Business Process Interdependencies Framework, a concise yet comprehensive description of what public health agencies do to prepare for, respond to, and recover from public health emergencies. The framework's 'monitor' group of processes includes *Conduct Syndromic Surveillance*, placing it in context of other emergency preparedness processes.

Methods

This session will present the framework, explain the methodologies used to develop it, and describe how syndromic surveillance and other surveillance processes interact with the rest of the framework. The framework's applicability will be demonstrated through two scenarios—a syphilis outbreak and a health agency's response to H1N1. We will also compare this framework with four existing emergency response frameworks—the National Response Framework,¹ the National Health Security Strategy,² the CDC goals for public health emergency preparedness (PHEP) activities, and the PREPARE for Pandemic Influenza Quality Improvement toolkit³—and discuss their relative advantages and uses.

We will show how this framework can accommodate local variations in structure, be used as a quality improvement tool, and aid in communicating public health's role to its emergency response partners.

Results

At the end of this session, participants will be able to:

1. Describe a framework that shows how public health agencies operate in public health emergency preparedness;
2. Understand the relative value of this framework in comparison with other emergency response frameworks; and
3. Describe how the disease surveillance process interacts with the other public health processes used in response to public health emergencies.

Conclusion

The framework generalizes well-to-health agencies nationwide, despite local variations in structure. Within the framework, the surveillance processes have an important role in triggering the initial response, as well as in informing continuing response activities. Public health leadership may use the framework to better organize, monitor, prioritize,

and fund public health preparedness activities. It can also be used as a quality improvement tool to understand and enhance preparedness processes for improved outcomes.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 National Response Framework. http://www.dhs.gov/files/programs/editorial_0566.shtm.
- 2 National Health Security Strategy. US Department of Health & Human Services/Assistant Secretary for Preparedness and Response; 2009:43.
- 3 Lotstein D, Leuschner KJ, Ricci KA, Ringel JS, Lurie N. PREPARE for Pandemic Influenza: A Quality Improvement Toolkit; RAND Health; 2008:217.

ABSTRACT

EHR-based syndromic surveillance during the 2009–2010 H1N1 pandemic

KS Fiske, M Siefert, and M Hoffman

Cerner® Corporation, Kansas City, MO, USA
E-mail: mhoffman1.cerner.com

Objective

This paper describes a national initiative connecting 850 non-affiliated healthcare provider organizations throughout the United States in order to provide situational awareness during the 2009–2010 H1N1 influenza A pandemic. We addressed the challenge of semantic variability between organizations through a centralized data-mapping approach.

Introduction

The global H1N1 influenza A pandemic in 2009 heightened the need for automated disease surveillance capabilities.

After an initial surge in confirmatory testing, clinicians moved to diagnosis based on patient assessment for fever combined with cough or sore throat, the influenza-like indicators (ILI). Although some organizations used automated data capture¹ or national systems with manual data entry (www.cdc.gov/flu/weekly/fluactivity.htm), there was not a turnkey national automated system in place to support syndromic surveillance for ILI among non-affiliated organizations. Semantic interoperability through standards utilization is widely expected to simplify large-scale data initiatives but is challenging with widely disparate uses of terminology.

Methods

Organizations utilizing Cerner electronic health record (EHR) systems were invited to participate in the HealthAware Flu Initiative. 166 organizations in 48 states representing 850 healthcare provider facilities are currently participating. Local terminologies from all participating organizations were reviewed by Cerner terminology curators. Data elements associated with ILI were mapped to a core vocabulary. Participating organizations installed an automated query that ran every 24 h and summarized the prevalence of patients with ILI, diagnostic test orders and positive influenza A results. These data, with limited demographical information and no personal health identifiers, were uploaded to the Cerner data operations team, aggregated and published to participating health organizations daily. Local (34), state (33) and national public health organizations (CDC) also received access to the system.

Results

The reference data models from the 166 organizations demonstrated widely disparate terminologies for the key ILI data elements. For example, 88 participating systems captured 'cough' using 101 potential prompt labels, with one organization using 17 different prompts. Fever was captured by 75 organizations using 72 potential prompts, whereas sore throat was captured by 26 organizations with only nine possible prompts. The responses to these prompts varied

widely. By 1 September 2009, 57 million patient encounters were monitored by this system for either syndromic or laboratory indications.

Conclusions

The HealthAware initiative implemented a large-scale nationwide network to monitor indicators of influenza. Participating organizations included ambulatory clinics, private hospitals, pediatric hospitals, academic medical centers and smaller regional medical centers. We found limited use of standard terminology amongst these organizations, highlighting the value of human-mediated curation until standards become more widely adopted.

Acknowledgements

This paper was presented as a poster presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

Reference

- 1 Schirmer P, Lucero C, Oda G, Lopez J, Holodniy M. Effective detection of the 2009 H1N1 influenza pandemic in the US Veterans Affairs medical centers using a national electronic biosurveillance system. *PLoS One* 2010;5:e9533.

ABSTRACT

Virtual public health informatics training: development and delivery of an online public health informatics course

Y Jiang, and CJ Staes

University of Utah, Salt Lake City, UT, USA
E-mail: yuling.jiang@utah.edu

Objective

This poster describes the development and delivery of an online American Medical Informatics Association (AMIA) 10 × 10 Public Health Informatics course at the University of Utah.

Introduction

Public health informatics is an emerging interdisciplinary field that uses information technology and informatics methods to meet public health goals. To achieve these goals, education and training of a new generation of public health informaticians is one of the essential components. AMIA's 10 × 10 program aims to realize the goal of training 10,000 health care professionals in applied health and medical informatics by the year 2010.¹ The Department of Biomedical Informatics of the University of Utah was established in 1964. As one of the largest biomedical informatics training programs in the world, the department is internationally recognized as a leader in biomedical informatics research and education.² The poster hereby describes the collaborative effort between Utah and AMIA to develop a public health informatics online course.

Methods

The course was developed based on the classroom version of the Public Health Informatics course taught in the Biomedical Informatics Department, with input from practitioners at the Utah Department of Health. The course contains six modules:

- Describe the mission and practice of public health and identify opportunities using informatics methods and tools;
- Describe fundamental informatics principles and their application to public health, including database design and process diagramming;

- Examine standards relevant to public health and create design artifacts to enable system interoperability;
- Describe the current and evolving relationship between clinical and public health systems;
- Examine roles required to develop and manage public health informatics projects and systems.

Each module contains a guidance file, one or more voice-over-power-point lectures, readings and resources, a short quiz and/or discussion. A student project was also developed for the students to look into a current public health information system of their own choice. An in-person session was planned at the end of course for the students to present their projects. Feedback was collected by distributing a student evaluation questionnaire to each student at the in-person session.

Results

The course was delivered from 18 January to 10 May 2010 using Blackboard Vista, the University of Utah's online

Figure 1 Average scores of students' evaluation of the Utah-AMIA 10 × 10 Public Health Informatics online course.

teaching resource. Fifteen students from Puerto Rico to California had various backgrounds from clinician, public health practitioner, to business development manager. Twelve students described surveillance workflows and information process at the in-person session before the AMIA NOW! conference. The students highly scored the course (Figure 1) and made comments such as: 'It covered a broad range of skills and knowledge relevant to public health informatics.' 'A new professional field, I would like to continue in this field (public health informatics).'

Conclusions

The Utah-AMIA 10 × 10 Public Health Informatics online course successfully introduced clinicians and public health practitioners to informatics principles and their

application to public health problems. With a high level of satisfaction, students developed basic analysis and lifelong-learning skills to engage in the evolving field of public health informatics.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 AMIA 10 × 10. Available at [https://www.amia.org/10 × 10](https://www.amia.org/10x10).
- 2 About the Department. Available at <http://www.bmi.utah.edu/?pageId=2182>.

ABSTRACT

Assessing address data quality for public health surveillance in Montreal

E Lanktree¹, C Jauvin^{1,2}, L Bédard¹, and D Buckeridge^{1,2}

¹Bureau de surveillance épidémiologique, Direction de santé publique, Agence de santé et des services sociaux de Montréal, Montréal, Canada; and ²Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Canada
E-mail: elanktre@santepub-mtl.qc.ca

Objective

To (1) validate an address verification algorithm (*Dracones qualité*, DQ, described by Zinszer *et al.*¹) developed to improve data quality for public health mapping and (2) identify the origin of address errors.

Introduction

In Montreal, notifiable diseases are reported to the Public Health Department (PHD). Of 44,250 disease notifications received in 2009, up to 25% had potential address errors. These can be introduced during transcription, handwriting interpretation and typing at various stages of the process, from patients, labs and/or physicians, and at the PHD. Reports received by the PHD are entered manually (initial entry) into a database. The archive personnel attempts to correct omissions by calling reporting laboratories or physicians. Investigators verify real addresses with patients or physicians for investigated episodes (40–60%).

The DQ address verification algorithm compares the number, street and postal code against the 2009 Canada Post database. If the reported address is not consistent with a valid address in the Canada Post database, DQ suggests a valid alternative address.

Methods

Individual disease episodes for Montreal residents reported to the PHD between 26 January and 9 May 2009 and followed by PH nurses were included ($n=798$). Exclusion criteria were H1N1 episodes, those whose investigator address was completely different from the initially reported address, invalidated or non-nominal episodes, for a sample of 408.

Complete investigator addresses treatable by DQ ($n=342$) were verified manually by a member of the research team (EL) using the Canada Post website.² These addresses were also scanned by DQ. DQ and Canada Post results were compared to determine whether DQ correctly identified valid and invalid addresses and proposed valid suggestions.

Calls were made to labs, hospitals or clinics, between 9 April and 9 July 2010, to confirm reported addresses. A list was compiled by selecting all cases that had a reported address deemed invalid by DQ, ($n=66$) and a 20% random sample³ of the 181 cases that had reports with hand-written addresses.

Results

DQ detected an error rate of 11.8% (47 of 398) upon initial data entry, 10.0% (40 of 401) at final entry, after archive verification, 18.0% from investigator addresses (60 of 333), and 36.5% from calls to reporting labs/physicians (23 of 63). All addresses corrected at data entry corresponded to DQ suggestions.

Of the 336 episodes with investigator addresses corresponding exactly to initial entry, 334 (99.4%) DQ suggestions correspond to Canada Post. For the two others, both numbered avenues (that is, ninth), DQ proposed one suggestion and Canada Post proposed two.

The investigator address was chosen as the gold standard, as it was closest to the case's actual address. Among initial entries, DQ detected three types of errors: 27 postal code, five street number and six street number or postal code errors. Most corresponding investigator addresses were identical. All other addresses (11 of 27, three of five, and one of six, respectively) that were corrected during the investigator calls corresponded to DQ suggestions.

Conclusions

The DQ algorithm is valid but short street names seem problematic. The algorithm will be refined using a street name length and edit distance solution.

Approximately two percent of errors detected at initial data entry were corrected through data entry and/or archive team efforts, but could be corrected by DQ if integrated into the electronic chart database, saving manpower and time. Labs and physicians should update patient addresses regularly to decrease error rates. Surveillance and intervention

would benefit as reliable data would improve disease cluster identification and communication with cases.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Zinszer K, Jauvin C, Verma A, Bedard L, Allard R, Schwartzman K, *et al.* Residential address errors in public health surveillance data: a description and analysis of the impact on geocoding. *Spat and Spatio-temporal Epidemiolo* 2010;1:163–8.
- 2 Canada Post Corporation 2010. <http://www.canadapost.ca/cpotools/apps/fpc/personal/findByCity?execution=e1s1>.
- 3 StataCorp 2009. STATA 10.1.

ABSTRACT

SAGES: a suite of freely available software tools for electronic disease surveillance in resource-limited settings

SL Lewis¹, BH Feighner¹, WA Loschen¹, RA Wojcik¹, JF Skora¹, JS Coberly¹, and DL Blazes²

¹The Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA; and ²Division of GEIS Operations, Armed Forces Health Surveillance Center, Silver Spring, MD, USA
E-mail: Sheri.Lewis@jhuapl.edu

Objective

This paper describes the development of the Suite for Automated Global bioSurveillance (SAGES), a collection of freely available software tools intended to enhance electronic disease surveillance in resource-limited settings around the world.

Introduction

Emerging and re-emerging infectious diseases are a serious threat to global public health.^{1,2} The World Health Organization (WHO) has identified more than 1100 epidemic events worldwide in the last 5 years alone.³ Recently, the emergence of the novel 2009 influenza A (H1N1) virus and the SARS coronavirus has demonstrated how rapidly pathogens can spread worldwide. This infectious disease threat, combined with a concern over man-made biological or chemical events, spurred WHO to update their International Health Regulations (IHR) in 2005.⁴ The new 2005 IHR, a legally binding instrument for all 194 WHO member countries, significantly expanded the scope of reportable conditions, and are intended to help prevent and respond to global public health threats. SAGES aims to improve local public health surveillance and IHR compliance, with particular emphasis on resource-limited settings.

Methods

More than a decade ago, in collaboration with the US Department of Defense (DoD), the Johns Hopkins University Applied Physics Laboratory (JHU/APL) developed the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE). The current SAGES initiative leverages the experience gained in the development of ESSENCE; the analysis and visualization components of SAGES are built with the same functionalities in

mind. Cognizant of work underway on individual surveillance systems components, for example, collection of data by cell phones, we have focused our efforts on the integration of inexpensive, interoperable software tools that facilitate regional public health collaborations.

Results

SAGES tools are organized into four categories: (1) data collection, (2) analysis and visualization, (3) communications, and (4) modeling/simulation/evaluation. Within each category, SAGES offers a variety of tools compatible with surveillance needs and different types or levels of information technology infrastructure. In addition to the flexibility of tool selection, there is flexibility in the sense that the analysis tools do not require a fixed database format. For example, rather than requiring an existing database to adapt to the tool, the SAGES database tools adapt to the format of all Java database compliant formats. Lastly, the SAGES tools are modular in nature, allowing the user to select one or more tools to enhance an existing surveillance system, or use the tools *en masse* for an end-to-end electronic disease surveillance capability. Thus, each locality can select tools from SAGES based upon their needs, capabilities, and existing systems to create a customized electronic disease surveillance system.

Conclusions

We have combined electronic disease surveillance tools developed at the Johns Hopkins University Applied Physics Laboratory with other freely available, interoperable software products to create SAGES. We believe this suite of tools will facilitate local electronic disease surveillance, regional public health collaborations, and international disease reporting. The Armed Forces Health Surveillance Center

welcomes inquiries on the SAGES tools from interested WHO member countries.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Binder S, Levitt AM, Sacks JJ, Hughes JM. Emerging infectious diseases: public health issues for the 21st century. *Science* 1999; **284**:1311–13.
- 2 Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. *Nature* 2004; **430**:242–9.
- 3 WHO. *The World Health Report 2007—A Safer Future: Global Public Health Security in the 21st Century*. World Health Organization: Geneva, 2007.
- 4 WHO. *International Health Regulations (2005)*. World Health Organization: Geneva, 2005.

ABSTRACT

Health IT meaningful use: how the EHR certification may improve the electronic laboratory reporting (ELR)

N Lipskiy, N Garrett, and S Tandon

Office of Surveillance, Epidemiology and Laboratory Science (OSELs), Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA
E-mail: dgz1@cdc.gov

Objective

The objective of this presentation is to evaluate the potential impact of Stage 1 meaningful use (MU) health IT certification (MUC), on development of national electronic laboratory reporting (ELR) capacities.

Introduction

The American Recovery and Reinvestment Act (ARRA) brought significant incentives to providers for implementing certified EHR technologies. It specifically requires utilization of certified electronic health records (EHRs) for electronic exchange of health information and for submission of clinical quality and other measures to the federal agencies. The most important barriers in the ELR implementation are a lack of funding at health departments, shortage of staff at health departments, and the variable content and format of ELR messages.¹ The MU is a new factor that may foster ELR technologies through implementation incentives and through standardization of EHRs.

Methods

(1) Analysis of ELR data structure and data exchange requirements (HL7) (a) for a communication between states and CDC programs and (b) across all MU objectives in comparison with the ELR public health reporting objective. (2) Impact assessment of EHR-related and ELR-related objectives on development of state and national ELR capacities. An assessment of states' ELR capacities was based on the 2008 ELR survey² and official web sites of states' department of health. We analyzed ELR messaging requirements using documents that were published through official HL7 and CDC web sites.

Results

A baseline assessment of ELR status before enacting of the ARRA demonstrates that most of the state legislations did not require ELR. The MUC fosters an implementation of ELR for notifiable conditions. Core MU objectives³ require an inclusion in certified EHRs data elements that describe patients'

demographics, clinical diagnoses, and test results. It aims to harmonize data elements in health IT systems that are relevant to ELR. For example, a MUC requirement on use of Logical Observation Identifiers and Codes (LOINC) for test procedures in hospital systems will foster using LOINC in ELR. The MUC of EHRs leverages national activities on creation, maintenance, and access relevant to ELR standards for content, messages exchange and their security, and patient privacy.

For example, because the MUC requires use of the HL7 version 2.5.1 for submitting of laboratory results to public health (PH) agencies, it may stimulate PH for using the same HL7 version for ELR. If an expected inclusion of certification requirements for a bi-directional communication between clinical care and PH is added to the MUC at Stage 2, then the MUC impact on ELR will be significantly increased.

Conclusions

Even though EHRs and ELR operate in different health IT systems (respectively, in clinical care and PH), the MUC of EHRs is a very important factor that may help in further ELR implementation, improvement of interoperability of state and national PH IT systems, and timeliness of PH emergency response.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Council of State and territorial Epidemiologists. About ELR. Available at: <http://www.cste.org/dnn/ProgramsandActivities/PublicHealthInformatics/tabid/346/Default.aspx>.
- 2 The 2008 National Electronic Laboratory Reporting (ELR) Snapshot Survey. Available at: <http://www.cste.org/dnn/Portals/0/2008NationalELRSurveySum.pdf>.
- 3 42 CFR Part 170, Health Information Technology: Initial Set of Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology, Final Rule, 2010.

ABSTRACT

Reporting nationally notifiable conditions (NNC): vocabulary aspects

S Ganesan¹, N Lipskiy², R Merrick³, and C Vinion

¹Northrop Grumman Consultant to CDC, Atlanta GA, USA; ²Office of Surveillance, Epidemiology and Laboratory Science (OSELs), Centers for Disease Control and Prevention (CDC), Atlanta GA, USA; and ³Independent Consultant to CDC, Atlanta GA, USA
E-mail: dsq3@cdc.gov

Objective

Electronic laboratory reporting (ELR) has a key role in public health case reporting and case notification. This paper will discuss the current status, problems, and solutions in a vocabulary support of nationally notifiable conditions (NNC) reporting.

Introduction

Standard vocabulary facilitates the routing and filtering of laboratory data to various public health programs. In 2008, Council of State and Territorial Epidemiologists (CSTE) developed 67 Technical Implementation Guides (TIGs) that accompany each condition and contain standard codes for NNC reporting. Those TIGs were reviewed by a public health subject matter expert panel (SMEP), in May 2010, consisting of members of the CDC CSTE Laboratory and PHIN Vocabulary and Messaging Communities of Practice Program, and representatives from the Regenstrief Institute and the International Health Terminology Standards Development Organization.

Methods

The SMEP reviewed CSTE position statements for NNC¹ and content of Table 2 of the 67 TIGs that were provided by CSTE. All laboratory criteria that exist in current position statements were compared with a list of general microbiology laboratory methods (that is, microscopy, antigen detection, antibody detection, DNA methods, and so on). Also, a list of existing Logical Observation Identifiers Names and Codes (LOINC) and Systematized Nomenclature of Medicine—Clinical Terms (SNOMED-CT) codes for notifiable conditions were compared with those that were listed in the TIGs. The Regenstrief LOINC Mapping Assistant, RELMA,² was used for retrieving LOINC codes sorted by each NNC and laboratory criterion. SNOMED-CT codes, as of May 2010, were used for a comparison with codes that were included in the TIGs.

Results

The SMEP concluded that laboratory vocabulary, after minor updates to current codes, in TIGs is ready to be implemented. The SMEP also suggested to CSTE a change in the format of the CSTE position statements. Specifically, the recommendations are to (1) add a section on 'Microorganisms/Agents' to applicable CSTE position statements,² (2) define a CSTE process for new development and maintenance TIG laboratory vocabulary, (3) clarify the use of paired serology tests, (4) add quantitative results in reports, and (5) develop additional value sets (a qualitative laboratory test finding value set based on SNOMED-CT evaluation finding domain and a specimen value set based on SNOMED-CT and HL7 as specified in the ELR v2.5.1 implementation guide).

Conclusions

There is a notable progression in the standardization of vocabulary for NNC reporting. The SMEP found existing problems (development of TIGs, vocabulary gaps, evaluation of laboratory results, and so on.) that may be eliminated by a defined process for collaborative work of public health and standard development organizations.

Acknowledgements

The authors would like to acknowledge the excellent participation of the CSTE Case Report Standardization Workgroup and the PHIN Laboratory Messaging Community of Practice. This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 CSTE. Position Statements at: <http://www.cste.org/dnn/AnnualConference/PositionStatements/09PositionStatements/tabid/333/Default.aspx> 2009.
- 2 Regenstrief LOINC Mapping Assistant, RELMA. Available at: <http://loinc.org/relma>.

ABSTRACT

Improving quality and interoperability of public health messaging guides: CDC messaging quality framework (MQF)

N Lipskiy¹, G Kesarinath¹, M Pray², and V Nwadiogbu²

¹Office of Surveillance, Epidemiology and Laboratory Science (OSELs), Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA; and ²Northrop Grumman Consultant to CDC, Atlanta, GA, USA
E-mail: dgz1@cdc.gov

Objective

This document describes the Public Health Information Network (PHIN) efforts on the development of the MQF, a flexible framework of services and utilities designed to assist public health partners with preparing and communicating quality, standard electronic messages.

Introduction

The HL7 messaging standard, version two¹ that was implemented by most vendors and public health agencies did not resolve all systems' interoperability problems. Design and tool implementation for automated machine-testing messages may resolve many of those problems. This task also has critical importance for rapid deployment of electronic public health systems.

Methods

(1) Assessment of messaging structures that support reporting of nationally notifiable conditions, NND² and Public Health Lab Interoperability Project, PHLIP.³ (2) Analysis of best practices on a structural validation of messages. (3) Analysis of CDC MQF⁴ capabilities for conformance testing and support rapid implementation of biosurveillance applications.

Results

Released in March 2010, a web-hosted MQF application (version 1.0) allows several options for loading messages. It contains a message validation tool based on a current messaging standard for a NND case notification,² which is constrained to a Generic case Notification Messaging Guide (MG) and MGs for Varicella and Tuberculosis. Another MQF option allows validating messages against a PHLIP MG for Influenza test reporting.³ At the first step of a message validation process the MQF application provides a structural validation of messages on message, segment, field and

component levels (that is, cardinality of segments groups, existence of required segments and their cardinality, existence and cardinality of segments and so on). The MQF then provides a message-constrained validation against messaging specifications (that is, validation of data types within OBX segments, lengths of specific data elements, verification that data elements are consist of only supported values and so on). The offline version adds a capability for synchronizing MQF with all NND MGs.

Conclusions

Even at the initial stage of development, the CDC MQF represents a flexible framework of services available to systems and public health partners inside and outside of CDC to prepare, test, and validate electronic messages against the relevant messaging, vocabulary standards, and program specific business rules.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Health Level Seven, Inc. *HL7 Messaging Standard Version 2.5: An Application Protocol for Data Exchange in Healthcare Environments*. Ann Arbor, MI, 2007.
- 2 PHIN HL7 Version 2.5. Messaging Standard National Condition Reporting Case Notification. ORU^R01 Message Structure Specification/Profile Version 2.0, October 23, 2008. Available at: <http://www.cdc.gov/phn/library/documents/pdf/PHIN%20Notification-MessageSpecificationProfile%20v2%200.pdf>.
- 3 APHL/CDC PHLIP Messaging Guide for Influenza Test Result Reporting by Public Health Laboratories, ORU R01 HL7 v2.3.1, Document version 1.0.2, Sept. 15, 2009. Available at: https://phinmqf.cdc.gov/HL7v2_3_1_APHL_Influenza_Msg_Guide_ORU_v1%200%202.pdf.
- 4 PHIN Message quality Framework. Available at: <https://phinmqf.cdc.gov/>.

ABSTRACT

Influenza messaging and interoperability: the PHLIP assistance team approach

J Lipsky¹, L Cohen¹, M Meigs², and P Zarccone²

¹The St John Group, LLC, Atlanta, GA, USA; and ²Association of Public Health Laboratories (APHL), Silver Spring, MD, USA
E-mail: jlipsky@TSJG.com

Objective

This paper describes the Public Health Laboratory Interoperability Project (PHLIP) assistance team (PAT) approach and the collaborative efforts between the Association of Public Health Laboratories (APHL) and the Centers for Disease Control and Prevention (CDC) to achieve electronic laboratory surveillance messaging (ELSM) for Influenza. The knowledge transfer and experience gained by state public health laboratories (PHLs) participating in PHLIP could serve as an interoperability model for other data messaging and surveillance initiatives.

Introduction

An essential theme of the US Federal Health Information Technology Strategic Plan is interoperability and the ability to effectively exchange information using specific data and technical standards.¹ In 2005, in an effort to accelerate the development of a national laboratory standards-based electronic data-sharing network, APHL and CDC collaborated to launch PHLIP.² The goals of PHLIP include, but are not limited to, improving the quality of data exchanged, piloting sustainable architecture for laboratory data exchange, sending and receiving HL7 test results from states to CDC programs (v2.3.1), increasing the use of Route-not-Read hubs for regional data exchange, and expanding these efforts beyond National Notifiable Diseases (NNDs). In an effort to achieve these goals, APHL solicited input directly from the PHL community to understand what assistance was necessary to achieve success with ELSM; in this case, Influenza as a prototype. After receiving feedback from PHLs responsible for reporting NNDs, the concept of technical assistance teams was formulated. In early 2010, APHL initiated an effort to send out the PATs to implement the ELSM message for Influenza in as many PHLs as possible by December 2010.

Methods

A detailed retrospective review of the experiences of the PATs experience will be presented. PHLs signed up for PAT support

and were prioritized according to input from the CDC's Influenza Division, as well as their own timeline for assistance. Deployed in spring 2010, two technical assistance teams were tasked with visiting PHLs to provide 'hands-on' assistance in areas including technical architecture, vocabulary, and project management expertise. States that were not available to host an onsite PAT visit were given the option of receiving 'virtual' support. Over the last 6 months, the process from initial state engagement until going 'live' with the PHLIP ELSM has evolved and best practices for achieving interoperable electronic data exchange have been captured.

Results

Presently, over 30 PHLs are working in some capacity on PHLIP, and 11 PHLs have been visited by PATs. Since March 2010, approximately 10 PHLs receiving either an onsite visit or virtual support by the PAT are now sending automated Influenza messages to CDC. Next states to be visited before December 2010 include NH, KY, NY, and WA, as well as others. (Table 1).

Conclusions

Over the last 6 months, PHLs have increased their interoperability functionality as well as enhanced their surge capacity for Influenza. These improvements are because of the PAT approach and the foresight of the collaborating organizations; APHL and CDC. Given the changing climate of infectious disease transmission, it is important to heed the lessons learned for effective laboratory information management processes, while keeping an eye toward interoperability in the public health domain.

Table 1 Process steps from initial PHL engagement until electronic message validation by state model type

Model	Time to validation step
Original, non-PAT states	6–12 Months
States with PAT support	Under a month

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Benson T. *Principles of Health Interoperability HL7 and SNOMED*. Springer-Verlag London Limited: London, UK, 2010.
- 2 Zarcone P, Nordenberg D, Meigs M, Merrick U, Jernigan D, Hinrichs SH. Community-driven standards-based electronic laboratory data-sharing networks. *Public Health Rep* 2010;**125** (2): 47–56.

ABSTRACT

Real-time, reusable, dynamic public health surveillance

P Mirhaji, R Vedam, and MM Vagnoni

School of Biomedical Informatics, University of Texas, Houston, TX, USA
E-mail: Parsa.Mirhaji@uth.tmc.edu

Objective

This paper describes use of semantic technologies in combination with Services Oriented Architecture (SOA) to construct dynamic public health surveillance systems¹ used for just-in-time monitoring of emerging infectious disease outbreaks. The system was used for surveillance of schools in the third largest population center, Harris County.

Introduction

The resources available in most public health departments are limited. Access to trained technical personnel and state-of-the-art computing resources are also lacking. Customizable off-the-shelf systems contribute only to creation of information silos, are expensive, and not affordable by the limited budget available to the departments of health (only growing worse with the recession). The one thing that has increased is the need for surveillance in more areas, from diseases to environmental exposures to unexpected disasters. One solution would be an adaptable system able to cope with changing requirements while reusing or eliminating infrastructure from both computing hardware and technical personnel.² We report in this paper an instance of such system as used to perform disease surveillance across the Harris County school system. The system is designed to be customizable for surveillance of any disease, while simultaneously accommodating other use cases like disaster response and registries.

Methods

The Survey-On-Demand System (SODS) enables epidemiologist to create all necessary data capture for an area to be surveilled. This eliminates the need for personnel to design a web page, desktop client, or the backend data store. In addition, the system ties directly into a surveillance dashboard that allows epidemiologist to view trends, set alerts, and monitor multiple surveillance projects simultaneously.

The system is designed to abstract the format of the data captured so that it can be seamlessly exported to other formats (that is, Excel or SAS). All the data captured is backed by controlled and/or colloquial terminologies, which enable

integration across multiple surveillance projects (past, present, and future) or information exchange with remote collaborators. Data is either mapped manually or natural language processing suggests appropriate terms from formal terminologies.

In order to cope with partial Internet connection (online somewhere, offline most places), an interaction model enables offline use and then synchronization when Internet is available. The server manages the state and macroscopic provenance of the forms. The client software logs every user interaction providing audit trails and item level rollback. All data is strongly encrypted. The system effectively manages the dynamic changes in the structure of the survey; no data is lost. The client keeps past data coherent with current data, as data collection needs change.

Results

More than 200 schools from Harris County were surveilled for absenteeism due to influenza-like illness through the academic year 2009–2010. The system is being evaluated to be used in other surveillance projects, including food-borne illness, environmental services, and children learning assessment. The SOA architecture enables a cloud-based information processing and data storage.³

Conclusions

The SODS system has been successful at both disaster and disease surveillance. It is being applied to multi-institutional clinical research and learning assessment in primary schools. The dynamic, model driven, information processing backend services support context independent, disparate use cases harnessing the same system and investment without creating another information/systems silo. The warehouse is a hybrid containing both relational tables for Online Analytics Processing (OLAP) and a semantic repository for flexible, extensible representation of highly variable data. In addition, the surveys can be shared across organizations and projects, enabling reuse and further minimizing resources required to do surveillance. There are many interesting possibilities to expand this surveillance toolkit.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Mirhaji P, Casscells SW, Srinivasan A, Kunapareddy N, Byrne S, Richards DM, *et al.* Services oriented architectures and rapid deployment of *ad-hoc* health surveillance systems: lessons from Katrina relief efforts. *AMIA Annu Symp Proc* 2006; 569–573.
- 2 Mirhaji P, Zhu M, Vagnoni M, Bernstam EV, Zhang J, Smith JW, *et al.* Ontology driven integration platform for clinical and translational research. *BMC Bioinformatics* 2009;10 (Suppl 2): S2.
- 3 Bechhofer SK, Stevens RD, Lord PW. GOHSE: ontology driven linking of biology resources. *Web Semantics: Science, Services and Agents on the World Wide Web* 2006;4:155–63.

ABSTRACT

HAISS Data Warehouse (HDW)—A new data access architecture for ESSENCE in the VA

A Mostaghimi¹, G Oda¹, C Lucero¹, P Schirmer¹, J Lombardo², R Wojcik², and M Holodniy¹

¹Department of Veterans Affairs, Office of Public Health Surveillance and Research, Palo Alto, CA, USA; and ²Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA
E-mail: anoshiravan.moshtaghimi@va.gov

Objective

To describe the new data warehouse, HAISS Data Warehouse (HDW) architecture whereby VA's Electronic Surveillance System for Early Notification of Community-based Epidemics (ESSENCE) will receive its required data elements from VA's 128 VistA systems in a more accurate, robust and time sensitive manner.¹

Introduction

The data elements required for the proper functionality of VA's ESSENCE system are all currently available within VA's 128 VistA systems. These data are made available to VA's ESSENCE system via a series of complicated MUMPS extraction routines, multiple data transformations crossing multiple servers, networks, operating systems and HL7-parsing

Figure 1 Required data elements are collected via a MDO-based ETL process and populate an ESSENCE data mart.

routines on a daily interval. With recent changes emerging in VA's information technology infrastructure, a new data architecture supporting ESSENCE's surveillance capabilities is becoming possible.

Methods

Healthcare Associated Infection and Influenza Surveillance System (HAISS) is a VA project tasked with the operation and deployment of VA's biosurveillance system using ESSENCE, and a nosocomial surveillance system. In order for HAISS to successfully untangle the current data access challenges supporting these systems, a new data access architecture has been developed and is being embedded as part of Extract-Transform-Load (ETL) layer of a data warehouse in support of HAISS Data Warehouse (HDW) project. The Extraction methodology is based on the Remote Procedure Call (RPC) capabilities of VistA. This VistA RPC technology is in turn abstracted into a Data Access Object (DAO) design pattern written in C# hence, an Application Programming Interface (API) called Medical Domain Objects (MDO) is written whereby targeted, inexpensive extraction calls are made against VistA systems. The resulting returning object in turn is passed along to a Business Processing Engine of the HDW whereby the Transformation-Load steps associated with populating the data warehouse and consequently the ESSENCE data mart are accomplished (Figure 1).

Conclusions

Utilization of MDO as part of the data extraction capabilities of HDW has offered us the ability to query for the required data elements using an extremely efficient and low overhead process. The required surveillance data are now made accessible to ESSENCE as queriable data inside a relational database hence, alleviating the expensive and often error-prone HL7-based flat-file parsing process. Finally, the Business Processing Engine associated with HDW allows the data warehouse manager the ability to poll the VistA data sources in a situationally specific interval, such that the periodicity of critical data elements used by ESSENCE-alerting mechanism is much shorter than a daily interval.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

Reference

- 1 Gillon, Joseph—VA SharePoint. <http://medora.sharepoint.med.va.gov/sites/mdws/default.aspx>.

ABSTRACT

Implementing an outbreak/event tracking monitor for public health

S Mottice¹, J Reid^{1,2}, T Jeppson¹, and R Rolfs¹

¹Utah Department of Health, Salt Lake City, UT, USA; and ²Department of Clinical Epidemiology, University of Utah, Salt Lake City, UT, USA
E-mail: smottice@utah.gov

Objective

The aim of this project was to create a secured web-based application that would run within the PH Access communication framework at the Utah Department of Health. This application—Epi Issue Tracker—would provide State and Local Health Departments with an environment that can be used to monitor actions during an event. We created a tracking system that functions as a combination dashboard/notification system to permit timely and effective communication of epidemiology events. This tracking system, Epi Issue Tracker, is used by all 12 Local Health Departments and the State Department of Health to share information across the state. There have been 830 issues/outbreaks entered into Epi Issue Tracker since January 2009, with 647 updates posted for those issues/outbreaks.

Introduction

Utah has a centralized State Health Department and 12 Local Health Departments situated throughout the state. Coordination of outbreaks or events that crosses jurisdictions has been historically difficult. Utah has not had a functional NEDSS-compliant database until 2009 and still does not have an Outbreak Management System (OMS). A survey was sent to Local Health Departments to assess their perception of need for real-time knowledge of current outbreak/events, with the majority indicating that current processes were inadequate.

Methods

We developed this application using open-source tools from PHAccess written in PHP and subject matter experts to provide initial application requirements and feedback. The development process used a series of iterative development cycles that included development, testing and feedback.

Through these cycles we are able to add product enhancements and fixes quickly and efficiently.

Results

We created Epi Issue Tracker and provided it to Local Health Departments during a sequential rollout. The application was designed to provide a dashboard detailing all ongoing events within the state. Events are individually managed and activities can be added and maintained in chronological sequence, thus expediting the creation of event reports. Completed events are moved to resolved status and can be maintained in a database that can be exported via CSV for tracking of outbreaks over time. There have been 830 events created since Epi Issue Tracker was rolled out to Local Health Departments, with 647 event updates posted to those events. Since April 2009 the system has been accessed 3099 times by both state and local epidemiologists.

Conclusion

This program is an easy-to-use issue tracker that facilitates cross-jurisdictional management of outbreaks and events.

Through Epi Issue Tracker, Local Health Departments are able to be informed of cross-jurisdictional outbreak events and post updates to those events for other Local Health Departments or the State Health Department. This application has helped increase communication within the state and improve relationships between the State and the Local Health Departments.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

ABSTRACT

Classification of errors for quality assurance in the emerging infections program influenza hospitalizations surveillance system

A Pérez, T D'Mello, L Kamimoto, L Finelli, and MH Torres-Urquidy

Influenza Division, Centers for Disease Control and Prevention, Atlanta, GA, USA
E-mail: hvv9@cdc.gov

Objective

Introducing data quality checks can be used to generate feedback that remediates and/or reduces error generation at the source.¹ In this report, we introduce a classification of errors generated as part of the data collection process for the Emerging Infections Program (EIP)'s Influenza Hospitalization Surveillance Project at the Centers for Disease Control and Prevention (CDC). We also describe a set of mechanisms intended to minimize and correct these errors via feedback, with the collection sites.

Introduction

The CDC's Emerging Infections Program monitors and studies many infectious diseases, including influenza.² In 10 states in the US, information is collected for hospitalized patients with laboratory-confirmed influenza. Data are extracted manually by EIP personnel at each site, stripped of personal identifiers and sent to the CDC. The anonymized data are received and reviewed for consistency at the CDC before they are incorporated into further analyses. This includes identifying errors, which are used for classification.

Methods

We evaluated the most current dataset as of 24 August 2010, containing records for 6521 persons with influenza-associated hospitalizations from 1 September 2009 through 30 April 2010. We built fully automated software routines using SAS version 9.2 (SAS Institute Inc., Cary, NC, USA) to conduct quality assurance. For instance: when data about the patient age are not provided, our software identifies the missing information as an error. We generated our classification based on the characteristics of these errors using a data-driven approach (that is, clustering errors with similar properties). The classification was then discussed internally. Based on the common characteristics of the clusters we developed common definitions for each category in the classification. Finally, we measured the actual number of errors in the most current collected dataset and categorized

Table 1 Error classification

Name	Predefined errors	Errors in reports	Ratio
Data entry errors	43	761	18.12
Missing data	19	263	13.84
Integrity	18	129	7.17
Failure to meet case definition	4	28	7.00
Chronology	32	113	5.35

the classification by type, with the most prominent ratio (predefined errors/errors in reports).

Results

The implementation of the error classification occurred during the preparation of the monthly report submitted to sites. The error classification was generated (Table 1). We found that 'Data Entry Errors' were the most prominent followed by 'Missing Data.' Other types of errors were identified as well.

Conclusions

Classification of errors allows for easier identification and prompt correction. In addition, it will allow us to improve subsequent versions of the software used to capture information and possibly minimize errors during capture.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Mullooly JP. The effects of data entry error: an analysis of partial verification. *Comput Biomed Res* 1990;23:259–67.
- 2 Dawood FS, Fiore A, Kamimoto L, Nowell M, Reingold A, Gershman K, et al. Influenza-associated pneumonia in children hospitalized with laboratory-confirmed influenza, 2003–2008. *Pediatr Infect Dis J* 2010;29:585–90.

ABSTRACT

BioSINE: an intuitive visualization tool to enhance collaboration between research and practice in disease surveillance

P Picciano¹, J Beaubien¹, G Mizrahi¹, J Owen¹, F Grant², and AM Hegedus²

¹Aptima, Inc., Woburn, MA, USA; and ²Northrop Grumman Corp., Atlanta, GA, USA
E-mail: ppicciano@aptima.com

Objective

BioSINE strives to improve situational awareness by making data visualization and collaboration capabilities intuitive and readily available for a wide range of public health (PH) stakeholders.

Introduction

Funded by the Army's Telemedicine and Advanced Technology Research Center (TATRC), we developed the BioSINE toolset to provide visualization and collaboration capabilities to improve the accessibility and utility of health surveillance data. Investigation of PH practitioners' needs with cognitive engineering methods revealed two key objectives:

1. To provide analysts and decision makers with an intuitive, visually driven workspace.
2. To support a web presence to provide rapid updating and facilitate greater interaction with data analysis in the PH community.

To better serve under-resourced PH organizations, both domestic and abroad, it is necessary to minimize information technology (IT) requirements and expertise in complex analytic tools.

BioSINE provides decision makers with the ability to create customized visualizations, focus on specific aspects of the data, or conduct hypothesis testing. Users can also view or hide variables, specify data ranges, and filter data relevant to their interests. Figure 1 shows a display in which a user investigated seasonal effects by narrowing the analysis to the summer months. Intuitive filtering is a key characteristic of the application to quickly produce snapshots of local interests.

Figure 1 Screenshot of consumer probing for seasonal trends by customizing date range.

Methods

We conducted usability tests with a sample of six participants, including three (military) public health experts and three (civilian) naïve users. Each performed a series of standardized tasks that required probing the dataset to identify morbidity patterns and answer specific questions. Morae[®] usability testing software was used to record the entire session, including the number of errors made, real-time verbal commentary, and survey items measuring satisfaction and usability opinions. Each questionnaire was scored on a five-point Likert scale with anchors ranging from *Strongly Disagree* (1) to *Strongly Agree* (5). After each testing session was completed, users participated in a semi-structured interview to suggest additional software features and functionality that might be useful to the larger PH community. Each user session lasted approximately 90 min.

Results

Users were quite successful in completing their tasks. Military and civilian users performed similarly. The mean number of errors per user was 4.17 (s.d.=2.04), none of which prevented completion. Users also reported they enjoyed using the BioSINE software. The mean satisfaction score was 4.07 (s.d.=0.59). Of the narrative comments collected, users reported that the BioSINE interface was easy to use, and provided a unique approach to interacting with and extracting information from health surveillance data.

Conclusions

Our findings over the course of this project suggest the ability to *interact with data*, and not just view reports,

leads to improved understanding, situational awareness, and collaboration. Our testing results indicate that even first-time users without PH expertise can successfully complete complex tasks and correctly answer health surveillance queries. Collected user opinions showed BioSINE to be useful and intuitive. Our intent is to harden the BioSINE prototype and make it a freely available web tool to the PH community.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

ABSTRACT

Mitigating data collection challenges with adaptive frameworks

A Poku¹, RA Wojcik¹, J Skora¹, S Lewis¹, and DL Blazes²

¹The Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA; and ²Division of GEIS Operations, Armed Forces Health Surveillance Center, Silver Spring, MD, USA
E-mail: adjoa.poku@jhuapl.edu

Objective

This paper describes the common challenges of data collection and presents a variety of adaptable frameworks that succeed in overcoming obstacles in applications of public health and electronic disease surveillance systems and/or processes, particularly in resource-limited settings.

Introduction

Electronic disease surveillance systems can be extremely valuable tools; however, a critical step in system implementation is collection of data. Without accurate and complete data, statistical anomalies that are detected hold little meaning. Many people who have established successful surveillance systems acknowledge the initial data collection process to be one of the most challenging aspects of system implementation.¹ These challenges manifest from varying degrees of economical, infrastructural, environmental, cultural, and political factors. Although some factors are not controllable, selecting a suitable collection framework can mitigate many of these obstacles. JHU/APL, with support from the Armed Forces Health Surveillance Center, has developed a suite of tools, Suite for Automated Global bioSurveillance (SAGES), that is adaptable for a particular deployment's environment and takes the above factors into account. These subsystems span communication systems such as telephone lines, mobile devices, internet applications, and desktop solutions—each has compelling advantages and disadvantages depending on the environment in which they are deployed. When these subsystems are appropriately configured and implemented, the data are collected with more accuracy and timeliness.

Methods

With SAGES we piloted multiple data collection methods against disease surveillance use cases. These pilots allowed

real-time assessment of initial requirements and evaluation of performance. The discussion describes various challenges encountered by users and implementers of data collection tools for disease surveillance systems: availability, privacy regulations, timeliness, latency, coverage, digitalization, automaticity, reliability, centralization, usability, sustainability, and cost.² Solutions for these obstacles are presented along with a comparative analysis of the advantages and disadvantages of the various frameworks. Field notes from data collection initiatives with collaborators in Peru (US Naval Medical Research Center Detachment) and the Republic of the Philippines (Philippines-Armed Forces Research Institute of Medical Services Virology Research Unit), as well as relevant analysis of popular frameworks, supplement the discussion and provide real-life examples. Emphasis is placed on a process of understanding the targeted locale and then adapting the collection framework to operate within that locale's environment and unique impacting factors to capture relevant, accurate, and timely information.

Results

The success of any data collection initiative is dependent on the type of framework put in place. The negative impact and long-term effects resulting from using inappropriate frameworks are explored during the discussion. Furthermore, the discussion defines a process for mitigating data collection challenges by leveraging adaptable tools (Figure 1).

Conclusions

With understanding of the targeted environment, the success of surveillance systems is achievable even with inherent challenges being present. The discussion offers a method for decomposing the environment, identifying potential obstacles, and comparing suitable solutions

Figure 1 Data collection quadrant of APL's SAGES suite is customizable for a deployment's needs.

for those challenges to define the data collection framework that will best meet the needs of stakeholders and end users.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Lewis SH, Wojcik R. Methodologies for data collection. *BMC Proc* 2008;2 (Suppl 3):S5. <http://www.biomedcentral.com/1753-6561/2/S3/S5>.
- 2 Ashar R, Lewis S, Blazes DL, Chretien JP. Applying information and communications technologies to collect health data from remote settings: a systematic assessment of current technologies. *J Biomed Inform* 2010;43:332–41.

ABSTRACT

The National Biosurveillance Integration System

H Park¹, E Pedersen¹, D Stetson¹, D Carr¹, D Bailey¹, M Stephens¹, J Best¹, A Marouf, and T Quitugua

National Biosurveillance Integration Center, Office of Health Affairs, Department of Homeland Security, Washington, DC, USA
E-mail: teresa.quitugua@dhs.gov

Objective

The National Biosurveillance Integration System (NBIS) is a consortium of federal agencies, whose joint objective is to enhance the identification, location, characterization, and tracking of biological events potentially impacting homeland security.¹⁻⁶ Together, the consortium members benefit from a joint awareness of potentially significant biological events that are unfolding or imminent, based on information shared among the group. This presentation describes the framework, activities and benefits for NBIS participants, and invites participation by other agencies.

Methods

NBIS integrates worldwide biosurveillance information across the domains of human health, animal and plant health, food and pharmaceutical issues, and relevant environmental factors. It leverages unique interagency communications and relationships to identify potential biological events of national concern. NBIS also supports prevention and mitigation of such events by providing timely notifications and ongoing situational awareness to enhance response. Biosurveillance information is gathered from a wide variety of government, scientific, restricted access, and open sources. Information regarding biological events and associated contextual information is then analyzed by NBIS subject matter experts who identify, characterize, and report events of concern in addition to their significance and potential impacts.

Results

Participating agencies report relevant information on the federal Biosurveillance Common Operating Picture (BCOP). The BCOP is a geo-temporal display of events of interest to the NBIS community, and is maintained by the Department of Homeland Security. The BCOP is currently accessible to

individuals within federal NBIS agencies. NBIS is developing a companion state, local, tribal, and territorial BCOP, which will share available information with non-federal government organizations. Expansion of this service to academic and private entities is under consideration.

Conclusions

NBIS and BCOP provide a robust information-sharing framework and communication pathway, facilitating timely and accurate notifications of biological events of concern to decision makers. This enhances the nation's ability to prevent, mitigate, and respond to significant events, efficiently and effectively.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, Utah, USA, on 1–2 December 2010.

References

- 1 Government Accountability Office (GAO)-10-171, Biosurveillance: Developing a Collaboration Strategy is Essential to Fostering Interagency Data and Resource Sharing. December 2009.
- 2 Government Accountability Office (GAO)-10-645, Biosurveillance: Efforts to Develop a National Biosurveillance Capability Need a National Strategy and a Designated Leader. June 2010.
- 3 Homeland Security Presidential Directive (HSPD)-9, Defense of United States Agriculture and Food, January 2004.
- 4 Homeland Security Presidential Directive (HSPD)-10, Biodefense for the 21st Century. June 2002.
- 5 Homeland Security Presidential Directive (HSPD)-21, Public Health and Medical Preparedness. October 2007.
- 6 Public Law 110.53, Implementing Recommendations of the 9/11 Commission Act of 2007, August 2007.

¹Contractors supporting the National Biosurveillance Integration Center.

ABSTRACT

Development of public health communication tools using open source methods

Jonathan Reid^{1,2}, Susan Mottice¹, Wu Xu¹, and Matthew Samore²

¹Utah Department of Health, University of Utah School of Medicine, Salt Lake City, UT, USA; and ²Departments of Internal Medicine and Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT, USA
E-mail: jreid@utah.gov

Objective

There is an urgent need for improved communication between stakeholders involved in outbreak investigations, public health reporting and events of interest occurring between different jurisdictions within the same state. Currently, state and local public health agency personnel rely on personal communications involving phone, fax and snail mail. The Utah Department of Health (UDOH) sought to develop and encourage the use of a secured web portal that allows access to a variety of applications using a single sign-on. This was achieved by developing a secured communications framework called PHAccess¹ that allows tools and applications to be implemented within a secure web environment, using open source software and Agile methodology techniques. The user-centric design currently hosts an electronic report-staging area, ELR/EMR reporting, web-based reporting, secure messaging between stakeholders and a state laboratory result look-up feature. Currently, there are over 700 registered users; 3693 secure messages that have been exchanged and the site has been accessed over 12,205 times since January 2009. Informal feedback from users has been encouraging and formal evaluation is planned, along with expansion and integration with state level health information exchange projects.

Introduction

Secure and confidential exchange of information is the cornerstone of public health practice. Often, this exchange has to occur between public health agencies across jurisdictions. Examples include notification of reportable diseases when the testing and residence of the patient are in different counties. The cross-jurisdictional issues become exaggerated in times of communicable disease outbreaks or events of interest that are not yet classified as outbreaks. Currently, such communication occurs between state and local agencies and between agencies and community clinicians

on a personal level, with phone, fax and snail mail. There are a multitude of secured websites hosted by UDOH that offer access to single applications requiring approved users to remember multiple sites and logins/passwords. The goal of this project was to develop a centralized, single sign-on secure web portal, from which users could access multiple applications and communicate securely with each other.

Method

The framework for PHAccess was developed using open source software, PHP and My SQL. The novel concept adopted by Apple iTunes App Store was used to develop our applications. This was supported by Agile methodology techniques with frequent iterative development cycles that were driven by constant feedback from users and stakeholders. A central premise was employing a user-centric design philosophy that allows users to influence the development of as well as manage and maintain the product.

Results

PHAccess was commissioned on 1 January 2009 after 4 months of development and iteration. The site started with three applications (Issue Tracker, Secure Messaging, Admin Functions) and has since grown to over 40. Since then, there are 744 registered users (54% state health, 14% local health departments and 32% community clinicians). The site has been accessed 12,205 times and 3693 secure messages have been exchanged. The current suite of applications include an EDI (Electronic Data Interchange) reporting suite of services that include electronic report-staging area, ELR/EMR reporting, web-based reporting, secure messaging between stakeholders and a state laboratory result look-up feature. There is also a query agent that allows the user to securely query and display results from external sources such as the Utah NEDSS or State laboratory information systems.

Conclusions

An integrated suite of services has been developed using open source tools and modest resources to provide public health agencies and clinicians a secure single sign-on access to relevant information. PHAccess is currently in use and further expansion is planned to integrate this service into state-wide clinical health information exchange projects.

Acknowledgements

This paper was presented a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

Reference

- 1 Program demonstration. <http://www.rockymountaincoe.org/phaccess>.

ABSTRACT

Identification of features for detection and prediction of homelessness from VA clinical documents

S Shen^{1,2}, BR South^{1,2}, M Palmer¹, S Duvall¹, M Samore^{1,2}, and AV Gundlapalli^{1,2}

¹*Salt Lake City VA Health Care System, Salt Lake City, UT, USA; and* ²*University of Utah, Salt Lake City, UT, USA*
E-mail: shuying.shen@hsc.utah.edu

Objectives

We demonstrate a semi-automated approach to induce and curate lexical domain knowledge for identification of evidence and risk factors for homelessness found in VA clinical documents. This domain knowledge can be used to support training and evaluation of automated methods such as Natural Language Processing (NLP) systems for detection and prediction of homelessness among veterans. This could serve as a proxy for public health and other surveillance involving homeless individuals. Similar methods could be used to identify other conditions of interest.¹

Introduction

Homelessness in general is a major issue in the US today. The risk factors of homelessness are myriad, including inadequate income, lack of affordable housing, mental health and substance abuse issues, lack of social support, and non-adherence to treatment/follow-up appointments. Early identification of these factors from clinical documents may help detect or even predict homelessness cases, allowing adequate intervention and prevention measures.

Methods

Using a think out loud approach, we developed an initial lexicon of features related to homelessness using expert inputs and available literature sources. This lexicon consists of social stressors (that is, recent divorce, unemployment), behavioral factors (that is, drug abuse), evidence (that is, lives in shelter, no housing), other risk factors (that is, exposure to war-related trauma) and direct mention of homelessness in the medical record (that is, homeless patient). This initial list was applied as pre-annotations to 600 VA clinical documents extracted from the VA Region one and four Data Warehouse for the time period 1/1/200-12/31/2009. Documents were pre-annotated using a prototype

system that supports interactive annotation and semi-automated curation of user-defined information classes.

Refining the Lexicon

Domain experts reviewed pre-annotated documents to determine if information was correctly identified, make modifications to annotations, add missing annotations, or reject annotations found to be incorrect or irrelevant. We applied an iterative process of revising the lexicon until further refinements were exhausted.

Results

Our initial lexicon had 83 entries. After two rounds of semi-automated curation on 75 documents, 38 concepts were added. Pre-annotations were helpful for reviewers to focus attention around the surrounding context, revealing important cues and textual patterns that would inform guideline development and creation of a reference standard for NLP system implementation.

Conclusions

Our approach can effectively generate lexical domain knowledge combining information from literature and expert feedback via iterative refinement. This method could be easily adapted to other surveillance efforts for case identification and prediction.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

Reference

- 1 Rosenheck R, Fontana A. A model of homelessness among male veterans of the Vietnam War generation. *Am J Psychiatry* 1994;151: 421–7.

ABSTRACT

Using clinician mental models to guide annotation of medically unexplained symptoms and syndromes found in VA clinical documents

BR South^{1,2}, M Palmer¹, S Shen^{1,2}, G Divita^{1,2}, S Duvall^{1,2}, M Samore^{1,2}, and AV Gundlapalli^{1,2}

¹Salt Lake City VA Health Care System, Salt Lake City, UT, USA; and ²University of Utah, Salt Lake City, UT, USA
E-mail: brett.south@hsc.utah.edu

Objectives

We sought to develop a guideline and annotation schema that can be consistently applied to identify medically unexplained syndromes (MUS) found in VA clinical documents. These efforts will support building a reference standard used for training and evaluation of a Natural Language Processing system developed for automated symptom extraction. Our overarching goal is to characterize the occurrence of MUS in Operation Enduring Freedom (OEF)/Operation Iraqi Freedom (OIF) veterans.

Introduction

Medically unexplained syndromes are conditions that are diagnosed on the basis of symptom constellations and are characterized by a lack of well-defined pathogenic pathways.¹ The three most common MUS are chronic fatigue syndrome (CFS), irritable bowel syndrome (IBS), and fibromyalgia. Different types of persistent symptoms, originating from different organ systems, characterize these syndromes. Patients often meet the criteria for more than one MUS.

Methods

Identification of MUS requires some level of inference on the part of annotators. For this reason we integrate clinician mental models with these review tasks. We demonstrate a practical approach that can be used to generate labeled data upon which additional layers of annotation can be added. Using this approach, annotated spans of text represent the most granular level of information necessary to identify symptoms and contextual features representing assertional information, and symptom duration. We developed an annotation guideline that provides definitions and examples of spans of text to annotate. Using a Protégé plugin called Knowtator,² we also developed an annotation schema used to mark spans of text representing symptoms and contextual features, along with the ability to link contextual features

with an annotated symptom. Once annotations have been created at the most granular level, additional assessments of medical explainability can be obtained from subsequent clinician review. We randomly sampled 492 clinical encounter documents from a cohort of OEF/OIF veterans, from the VA Region one and four Data Warehouse, who received health care services between 1 January 2007 and 12 December 2010. As part of tasks related to refining our annotation guideline, four reviewers annotated a random sample of 15 of these documents using the annotation guideline and schema. We report the estimated number of annotated symptoms and the number of symptoms that could potentially be included in non-mutually exclusive symptom constellations for the three most common MUS.

Results

Preliminary testing shows promising results for these methods. The number of words in the 15 annotated documents varied from 554 to 3314, with an average of 17 symptom annotations per document. The number of annotations (unique mentions) for all four annotators was 1358 (583) overall with symptoms 1010 (477), followed by assertions 262 (53), and symptom duration 86 (53). Clinician *post hoc* review of the 477 unique symptom annotations revealed non-mutually exclusive symptom clusters of which 274 (57%) could potentially describe CFS, IBS 29 (6%), and 155 (32%) fibromyalgia.

Conclusions

Our methods demonstrate a practical way of conducting annotation tasks that allow reuse of labels for other symptom surveillance efforts. Once symptoms have been annotated, additional layers of review can be done by another group of clinician annotators. Annotated data can be used to train NLP tools for related surveillance tasks including extracting symptoms representing known and

unknown syndromes of interest to clinicians, epidemiologists and, public health agencies.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Fink P, Toft T, Hansen MS, Ørnbøl E, Olesen F. Symptoms and syndromes of bodily distress: an exploratory study of 978 internal medical, neurological, and primary care patients. *Psychosom Med* 2007;**69**:30–9.
- 2 Ogren PV. Knowtator: a Protégé Plug-in for Annotated Corpus Construction. *Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*. Association for Computational Linguistics: Morristown, NJ, USA, 2006. pp. 273–275.

ABSTRACT

Document classification toward efficient event-based biosurveillance

M Torii, N Nelson¹, and D Hartley¹

Division of Integrated Biodefense, ISIS Center, Georgetown University Medical Center, Washington, DC, USA
E-mail: torii@isis.georgetown.edu

Objective

This paper describes ongoing efforts in enhancing automated document classification toward efficient event-based biosurveillance.

Introduction

Event-based biosurveillance is a practice of monitoring diverse information sources for the detection of events pertaining to human health.^{1–3} Online documents, such as news articles on the Internet, have commonly been the primary information sources in event-based biosurveillance.^{4–8} With the large number of online publications as well as with the language diversity, thorough monitoring of online documents is challenging. Automated document classification is an important step toward efficient event-based biosurveillance. In Project Argus, a biosurveillance program hosted at Georgetown University Medical Center, supervised and unsupervised approaches to document classification are considered for event-based biosurveillance.

Methods

In Argus operations, analysts are requested to label online documents that they read in their regular surveillance work. Currently, two document classes, relevant and irrelevant, are assumed. With such labeled articles, a customized classifier is trained for target geographic regions/languages using a machine-learning algorithm. Documents retrieved from a Boolean keyword search can be classified (filtered) or ranked according to the relevancy scores assigned. In addition, we considered dynamic grouping of documents, in contrast to classification into predefined classes. To reflect analysts' perspective in clustering documents, we try to weight features (for example, keywords) based on information extracted from class-labeled documents and/or past event reports.

Results

We have tested the proposed framework to facilitate supervised machine-learning classifiers on past data. The framework has been implemented in the Argus surveillance workflow. We are in the process of evaluating the performance of trained classifiers in operational settings. Meanwhile, we became aware of inherent challenges in the framework that could affect performance of classifiers, which include class-imbalance in training data sets, that is, few labeled irrelevant (or relevant) articles may be available because of labeling bias (or to the inherent class-imbalance), and dominant topics in labeled examples, for example, articles on seasonal influenza.

Using past data, we observed that informative subsets could be derived using document clustering, for example, k-means clustering.

Conclusions

We observed promising results on automated document classification in our preliminary experiments. Previously good results have been reported by other related studies in this domain.^{4–8} Meanwhile, Boolean queries created and maintained by expert analysts have also been found effective in Project Argus. The utility of automated document classifiers in contrast to the Boolean keyword search should be evaluated in real-life surveillance settings in the future.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Hartley DM, Nelson NP, Walters R, Arthur R, Yangarber R, Madoff L, *et al.* The landscape of international event-based biosurveillance. *Emerg Health Threats J* 2010;3:e3.

¹These authors contributed equally to this paper.

- 2 Keller M, Blench M, Tolentino H, Freifeld CC, Mandl KD, Mawudeku A, *et al.* Use of unstructured event-based reports for global infectious disease surveillance. *Emerg Infect Dis* 2009;15:689–95.
- 3 Walters R, Harlan P, Nelson NP, Hartley DM. Data sources for biosurveillance. In: Voeller JG (ed) *Wiley Handbook of Science and Technology for Homeland Security: Risk Analysis*. Hoboken, NJ, USA, 2009.
- 4 Nelson NP, Brownstein JS, Hartley DM. Event-based biosurveillance of respiratory disease in Mexico, 2007–2009: connection to the 2009 influenza A(H1N1) pandemic? *Euro Surveill* 2010;15, pii=19626. Available online <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19626>.
- 5 Brownstein JS, Freifeld CC. HealthMap: the development of automated real-time internet surveillance for epidemic intelligence. *Euro Surveill* 2007;12:E071129–071125.
- 6 Mawudeku A. Mining the Internet: GPHIN. In *The International Meeting on Emerging Diseases and Surveillance (IMED 2007)*. Vienna, Austria, 2007.
- 7 Collier N, Doan S, Kawazoe A, Goodwin RM, Conway M, Tateno Y, *et al.* BioCaster: detecting public health rumors with a Web-based text mining system. *Bioinformatics* 2008;24: 2940–1.
- 8 Steinberger R, Fuat F, Groot Evd, Best C, Etter Pv, Yangarber R. *Text Mining from the Web for Medical Intelligence*. OIS Press, The Netherlands, 2008.

ABSTRACT

Building an automated Bayesian case detection system

F-C Tsui, J Espino, T Sriburadej, H Su, and J Dowling

RODS Laboratory, Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
E-mail: tsui2@pitt.edu

Objective

This paper describes the architecture and evaluation of our recently developed automated Bayesian case detection (BCD) system.

Introduction

Current practices of automated case detection fall into the extremes of diagnostic accuracy and timeliness. In regards to diagnostic accuracy, electronic laboratory reporting (ELR) is at one extreme and syndromic surveillance is at the other. In regards to timeliness, syndromic surveillance can be immediate, and ELR is delayed 7 days from initial patient visit.¹

A plausible solution, a middle way, to the extremes of diagnostic precision and timeliness in current case detection practices is an automated Bayesian diagnostic system that uses all available data types, for example, freetext ED reports, radiology reports, and laboratory reports. We have built such a solution—BCD. As a probabilistic system, BCD operates across the spectrum of diagnostic accuracy, that is, it outputs the degree of certainty for every diagnosis. In addition, BCD incorporates multiple data types as they appear during the course of a patient encounter or lifetime, with no degradation in the ability to perform diagnosis.

Methods

The BCD system that we built has five components: real-time HL7 parsers, natural language processing (NLP) tools, Bayesian inference engine, a Bayesian network, and a database. The HL7 message parsers extract different data types from HL7 messages. Then, NLP tools, MedLEE,² and Topaz (a homegrown tool) find medical terms contained in each freetext report, including significant negative findings. We store the NLP results in a database. For non-freetext reports such as laboratory reports, we store the coded data directly to the database.

We built a Bayesian network with 57 nodes for detection of three diseases: flu, shigellosis, and measles. We formed the network structure and conditional probabilities by consulting a physician board certified in infectious diseases. We implemented the Bayesian inference engine in Java using the junction-tree algorithm. To make the BN portable, each node

in the network is represented in an Unified Medical Language System (UMLS) Concept Unique Identifier.

We performed a preliminary evaluation of the BCD system using only freetext ED reports as input to detect influenza cases. The gold standard was laboratory-confirmed positive and negative reports.

Results

Our preliminary evaluation used 363 randomly selected reports (181 positive) from 12 January 2005 to 31 August 2007. We found an area under ROC curve of about 0.8 (95%CI: 0.76–0.85). When the posterior probability threshold was set to 0.8 (given $P(\text{flu}) = 0.1$), we found a sensitivity of 63.5% (95%CI: 56.5–70.5%), a specificity of 82.4% (95%CI: 76.9–87.9%), and a positive predictive value of 78.2% (95%CI: 71.5–84.9%).

We used BCD to estimate daily expected counts of flu cases presented in seven EDs of UPMC health system during H1N1 outbreak by summing the posterior probability for flu for each visit. Figure 1 shows chart of percent daily expected ED flu visits from July to December of 2009. Average daily ED visits was 569.

A demo web page of BCD is available at <https://betaweb.rods.pitt.edu/casedetection-rest/demoPage.jsp>. It demonstrates multiple data types' input for computing posterior probabilities of three diseases.

Conclusions

Our BCD system has good performance characteristics and is a solution to low diagnostic accuracy and timeliness in existing automated surveillance systems.

Figure 1 A flu chart showing daily percentage of ED visits with flu (green) and its 5-day moving average (black) between 7/1/2009 and 12/1/2009.

Acknowledgements

The project is funded by CDC grants P01 HK000086 and 1U38 HK000063-01. This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Que J, Tsui FC, Wagner MM. Timeliness study of radiology and microbiology reports in a healthcare system for biosurveillance. *AMIA Annu Symp Proc* 2006; 1068.
- 2 Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc* 2004;11:392–402.

ABSTRACT

Epibasket: a prototype web tool aimed to speed up and standardize the epidemiological investigation of an outbreak

W Xing¹, G Hejblum¹, GM Leung², and A-J Valleron¹

¹Ecole Pierre Louis de Santé Publique, Université P. & M. Curie & Inserm, Paris, France; and ²School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China
E-mail: Alain-jacques.valleron@upmc.fr

Objective

This paper describes how the ideas and tools of e-commerce can be translated to the investigation of outbreaks: epidemiologists will 'shop' the best available items for their questionnaire, enhance the chances of producing interoperable questionnaires, and speed up the whole process.

Introduction

The detailed analysis of the epidemiological literature on the 2003 SARS epidemic published in peer reviewed journals has shown that a majority (78%) of the epidemiological articles were submitted after the epidemic had ended, although the corresponding studies had relevance to public health authorities during the epidemic. The conclusion was that to minimize the lag between research and the exigency of public health practice in the future, researchers should consider adopting common, predefined protocols and ready-to-use instruments to improve timeliness, and thus, relevance, in addition to standardizing comparability across studies.¹

Methods

The basic idea of the method is to avoid the epidemiologists faced to an outbreak to reinvent the wheel, and to provide them an online 'catalogue' of the 'best' items to include in their questionnaire. At the present time, the prototype concerns the psychobehavioral studies performed during an outbreak. We have shown¹ that they accounted for 19% of the studies made during the 2003 SARS epidemic. Questionnaires concern knowledge, attitudes, behaviors, and psychological and social impacts. The catalogue of possible questions was constituted from the comprehensive review of the relevant papers ($n = 59$) published during the SARS epidemic. The 1962 variables collected in these papers were standardized with the aid of usual medical terminolo-

gies (for example, MeSH, UMLS, SNOMED, CDISC standards and so on). The user interface was built by adapting a freely available software used for e-commerce (Prestashop²). The queXML tool² was used to generate the questionnaire with the selected items in pdf format, and/or to generate a file that could be used with online survey development tools (for example, LimeSurvey, an open source free software²) to create an online questionnaire.

Results

In the present prototype, the epidemiologist 'shops' the items and puts them in a 'shopping basket'. At the end, he gets a paper questionnaire or a file of questions. Each item is documented (Figure 1). The parallel with e-shopping can be pursued to provide a 'value' for each item (for example, the

Figure 1 Description of an item from the online catalogue (users get the definition, MESH terms, and list of papers having used it).

value might include the Impact Factor of the journals that have published the papers having previously used the variable). This may be useful, for example, at the time of choosing between two multidimensional scales of risk perception.

Conclusions

The concept of EpiBasket could increase the rapidity of the data collection in the case of an outbreak, and help to standardize the data collected. Moreover, the constitution of a catalogue of 'important' items for investigative epidemiologists during an outbreak could be a collective work of ISDS.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

- 1 Xing W, Hejblum G, Leung GM, Valleron AJ. Anatomy of the epidemiological literature on the 2003 SARS outbreaks in Hong Kong and Toronto: a time-stratified review. *PLoS Med* 2010;7:e1000272.
- 2 See www.prestashop.com, quexml.sourceforge.net/, and www.limesurvey.org.

ABSTRACT

Map application to the New York state electronic syndromic surveillance system

C Weng, J-H Chen, C DiDonato, and H-G Chang

Statistical Unit, Division of Epidemiology, New York State Department of Health, Albany, NY, USA
E-mail: cxw14@health.state.ny.us

Objective

The objective of this paper is to describe a map application added to the New York state Electronic Syndromic Surveillance system (ESSS). The application allows system users to display the geographic distributions, and trends of fever syndrome that was used to monitor seasonal and H1N1 influenza activities.

Introduction

The primary goal of the Electronic Syndromic Surveillance system (ESSS) is to monitor trends in non-specific symptoms of illness at the community level in real time. The ESSS includes emergency department chief complaint data that are categorized into eight syndromes: respiratory, gastro-

intestinal, fever, asthma, neurological, rash, carbon monoxide, and hypothermia. Since the onset of H1N1, fever syndrome has been used to monitor flu activity. As H1N1 spread nationwide, the need of visualizing flu activity geographically became clear, and urgent.

Methods

SAS software (SAS Institute Inc., Cary, NC, USA) is used to generate the maps. Once the users log on to the ESSS, enter map-generating parameters via an HTML screen, a Java-based web application is developed to run SAS codes to produce the maps. As this is a web application, the backend systems will respond with the data back to the client within five minutes, which allows for on demand display of the resultant maps.

Figure 1 A snapshot of the user-generated map with pop-up label.

Figure 2 A snapshot of the statewide fever syndrome map.

To reduce the SAS running time for the end users, a separate SAS program was scheduled to run every morning before business hours that creates a small-sized sub-dataset for query. The SAS output is constrained to HTML and GIF images to meet the system requirement.

Results

The map application was implemented in August 2010. It contains two components. The first component is a user-generated geographic syndrome display. With this component, users input three parameters, which are where (that is, the geographic level by county, or counties), when (that is, one day in the past seven days that the users want to study), and what to generate (fever syndrome). On each resultant map, counts of emergency department fever syndrome visits are aggregated, and displayed as a rate (per 100,000 population) by patients' ZIP Codes. The ZIP Code population is represented with color shading in the resultant maps; the darker the color, the larger the population in that ZIP Code. The fever syndrome visit rates are denoted by three levels of colored dots. In addition, if users hover their cursor over a ZIP Code, a pop-up label will also appear, identifying the ZIP Code, county name, number of emergency department fever syndrome visits, population in that ZIP Code, and the fever syndrome visit rate (Figure 1). The second component of the application is a weekly statewide display of the fever syndrome visit counts. When requested, statewide fever syndrome maps of the most current four weeks will display

to show the trend. Each map represents one of the CDC Morbidity and Mortality Weekly Report (MMWR) week. On the maps, emergency department fever syndrome visits are aggregated by patients' ZIP Codes. The maps are shaded by fever syndrome visit counts; the darker the color, the higher the fever syndrome visit counts (Figure 2). Users can also zoom in the maps or view the historical weekly maps for the past twelve months.

Conclusions

The map application is a new data visualization tool for the ESSS users. It provides hospital, local health department, region, and state users a user-friendly platform to monitor H1N1 and seasonal flu activities of their own jurisdiction, and the neighboring areas. The two components of the application enable users to generate informative local maps of fever syndrome visits, as well as to see the statewide trends over time. The map application does not require user-end software, or risk data security. It can be easily adapted for surveillance of other syndromic indicators.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

ABSTRACT

Enhance public health data capacity through statewide health information exchange of immunization records

W Xu, F Lanier, and N McConnell

Utah Department of Health, Salt Lake City, UT, USA
E-mail: wxu@utah.gov

Objective

The goal of the immunization information exchange is to expand the Utah Statewide Immunization Information System (USIIS) interoperability to all private providers, especially those who are not a USIIS' user but participate in clinical Health Information Exchange (cHIE).

Introduction

Utah Statewide Immunization Information System (USIIS) is the state immunization registry. It is connected to about 700 organizations, including 100% public health clinics, 60% of private providers, many schools, daycares, pharmacies, and Indian Health Services. Data exchange methods range from web data entry, batch file transfer, proprietary or HL7 data interfaces, and real-time exchange with Intermountain Healthcare's electronic health records (EHR). Clinicians in Utah ranked immunization data as one of the first five use cases for developing the statewide clinical Health Information Exchange (cHIE) in Utah Health Information Network. Utah Department of Health (UDOH) has collaborated with Utah Health Information Network to develop the immunization information exchange.

Method

We used the agile project management method and developed a roadmap with two developmental tracks and two implementation phases. The two tracks are parallel efforts to develop (1) policies and standards, and (2) technical infrastructure. The first phase is to 'push' the USIIS consolidated immunization records to the cHIE Virtual Health Records (VHR) for provider to query. The second phase aims to 'pull' new immunization information from the cHIE Virtual Health Records into USIIS.

Results

The policy track has completed following governance milestones: *The Electronic Commerce Agreement* is signed among all

Utah cHIE partners. This is a statewide agreement covering all Utah Department of Health (UDOH) programs' exchange with cHIE providers.

Utah Department of Health enterprise-wide management policy and procedures are developed on the basis of USIIS and Medicaid regulations and management protocols. Roles and responsibilities for operating and maintaining the interoperability among independent information systems and cHIE are defined.

HL7 2.5.1 Immunization Standards with CDC CVX codes are formally adopted for cHIE. USIIS has worked with UHIN Standards Committee to complete the implementation specifications for cHIE users.

The technical track for phase I has developed or is developing:

The UDOH-HIE gateway

This gateway will serve all public health clinical information exchanges with cHIE. USIIS has loaded records to cHIE, including demographics, consolidated immunizations, contraindication, and exemption information. Weekly updates capacity is established as well.

CVX codes mapping

Some EHRs use national drug codes (NDC) codes to store vaccine records in drug tables. According to the standards, cHIE is mapping USIIS' CVX codes with NDC codes in the cHIE Virtual Health Records.

Displaying consolidated immunization information in patient summary has been the cHIE's latest, hopefully the last challenge to accurately present immunization records from multiple sources for cHIE users.

Conclusion

New policy and IT infrastructure are needed to enable public health to exchange clinical information through statewide HIEs. Immunization registry can be a good pilot to develop a roadmap for this endeavor.

Acknowledgements

Acknowledgement to the CDC Rocky Mountain Center for Translational Research on Public Health Informatics (PI Matt Samore) and ONC funds for State HIE Cooperative

Agreement Program (State HIT Coordinator David N Sundwall). This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

ABSTRACT

Public health data sharing policy and informatics initiatives at China CDC

Y Zhang, and Q Chen

Chinese Center for Disease Control and Prevention (China CDC), Beijing, Peoples Republic of China
E-mail: yingjie924@yahoo.com.cn

Objective

The objective of this study is to describe public health data-sharing policy, and informatics initiatives at China Center for Disease Control and Prevention (China CDC).

Introduction

Construction of data-sharing network for public health is one of the national scientific data-sharing projects, based on the data resource that distributed at China Center for Disease Control and Prevention (China CDC), universities, research institutes, and scientists, as well as the data from research projects were integrated.

Data sharing policy

China CDC believes that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health. The policy reaffirmed the principle that data should be made as widely, and freely available as possible while safeguarding the privacy of research participants, and protecting confidential and proprietary data.

Informatics initiatives

From 2004, China CDC has launched many informatics initiatives, including: to establish the technical platform

releasing to the entire society, to push forward China's scientific data sharing, to assemble more valuable and available databases especially for epidemiology of the serious diseases, to build database integration for special research field in order to give more convenience to users, to set up the rules for technical standards data sharing, management, quality control of the databases, and service guide. The goal of these initiatives is to build infrastructure and networks to facilitate data sharing, integration, and interoperability.

Conclusions

Whole public health information, when combined with clinical and other phenotype data, offers the potential for increased understanding of disease processes affecting human health, improvement in the disease control, and prevention. For these reasons, China CDC believes that the full value of infrastructure and networks can be realized only if the public health datasets are made available as rapidly as possible to a wide range of scientific investigators.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.