A Protégé Ontology as The Core Component of a BioSense Message Analysis Framework

Cecil Lynch\textsuperscript{1,2}, Craig Cunningham\textsuperscript{1}, Eric Schripsema\textsuperscript{1}, Tim Morris\textsuperscript{3}, Barry Rhodes\textsuperscript{3}

\textsuperscript{1}OntoReason, LLC, \textsuperscript{2}UC Davis, \textsuperscript{3}US Centers for Disease Control and Prevention
Outline

• BioSense description
• Describe the current environment
• Describe the ontology
• Describe the ontology framework
• Describe the analysis workbench
• Future directions
• Questions
BioSense Description
What is BioSense?

- Real-time and near real-time national public health message analysis framework
- Consists of
  - Message acquisition and translation interfaces
  - Secure message transmission network
  - Message classification components
  - Data storage and query components
  - Data analysis component
  - CDC Monitors
  - Local data visualization and distribution
**BioSense Functions**

Confirm or refute existence of an event
- Environmental signal
- Suspect illness
- Intelligence warning
- Known outbreak/public health event

Monitor ongoing event and effectiveness of response
- Ascertain size of event
- Ascertain rate of spread
- Track efficacy of response efforts
- Monitor for adverse events
- Know when an event has passed
### Data Sources

<table>
<thead>
<tr>
<th>Data Source</th>
<th>2006</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orders &amp; results from 3 major commercial clinical laboratories</td>
<td></td>
<td>Represent 20% of all US lab testing; 60% of independent testing; critical to many PH efforts</td>
</tr>
<tr>
<td>Real-time data from VA</td>
<td></td>
<td>150 hospitals and ~1000 ambulatory care clinics; share data with many state and local PH communities</td>
</tr>
<tr>
<td>Real-time data from DoD</td>
<td></td>
<td>45 US hospitals and ~800 ambulatory; share data</td>
</tr>
<tr>
<td>Poison Control Centers call data</td>
<td></td>
<td>All 62 poison control centers; display and compare with other community health data</td>
</tr>
<tr>
<td>Private Hospitals</td>
<td></td>
<td>500 Clinical care Hospitals provide national view and local data</td>
</tr>
</tbody>
</table>
Target Data Types

- **Foundational**: demographics, chief complaint, discharge diagnoses, disposition, hospital utilization
- **Clinical**: vitals, triage notes, working diagnosis, discharge summary
- **Laboratory**: orders, microbiology results
- **Pharmacy**: medication orders
- **Radiology**: orders, interpretation results

All structured in HL7 2.5 BioSense messages
Current Classification

• Data mapped to 11 syndrome categories
  • Botulism-like
  • Fever
  • Gastrointestinal
  • Hemorrhagic illness
  • Localized cutaneous lesion
  • Lymphadenitis
  • Neurological
  • Rash
  • Respiratory
  • Severe illness/death
  • Specific infection
• 79 sub-syndrome categories
Watch what you ask for!

- BioSense message volume capacity today
  - 837 messages a second
  - >72 million messages a day
- How does an epidemiologist review that volume of data?
- How do you link messages to an individual over time to refine the diagnostic info?
Current BioSense Framework
Load Balancing
Message Type Filter

The JMS Input Queue will be deployed across a cluster.

A Java Servlet will be used in the future to assign priorities to specified message types.

An Enterprise Java Bean will write the raw batches to a database. If the batch is MIXED, it will de-batch it and route the messages to the appropriate JMS queues. If it is a single type, it will simply route it to the appropriate queue.

Each JMS Type Queue will support a single message type. For each message, it will call the onMessage() handler of the first available Rhapsody JMS listener for that message type.

Each Rhapsody server will have a JMS Listener for each message type. It will connect to the calling queue, retrieve the message, validate it, transform it, and publish it to the appropriate topic or queue.

DMB — Data Messaging and Brokering
ETL Processing
AV and OTP
End User Views

BioSense Users
On
The World Wide Web
### Concept In HL7 V3 Data Type

<table>
<thead>
<tr>
<th>Code</th>
<th>CodeSystemVersion</th>
<th>CodeSystemName</th>
<th>Qualifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>49727002</td>
<td>0601core</td>
<td>01/31/2006</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SNOMED CT</td>
<td></td>
</tr>
</tbody>
</table>

**Term:** Cough

**Code:** 49727002

**CodeSystemVersion:** 0601core

**CodeSystemName:** SNOMED CT

**DateChanged:**

**OriginalText:** Cough

**Translation:**
- Cough (finding)
- Observation of cough
- Finding of cough
- Cough

**Name:** Lynch82824016SN

**BioSenseMatch:**
- toss 1221
toss 1221
- couhg 300
- 7862 3902
- Cough 3298
- Cough with fever

**Systems and synonyms:**
- Decreased diffusion capacity
- Increased peak expiratory flow
- Increased oxygen supply
- Asthma finding
- Pulmonary sign AND/OR

**Children:**
- Decreased coughing
- Bovine cough
- Nocturnal cough
- Early morning cough
- Spasmodic cough
- Hemoptysis
- Chronic cough
- Cough at rest
- Cough suppression
- Postural cough
- Dry cough
- Cough with fever

**Parent:** 

**Other code:**
- BioSense Terms

**Parent:** Systems and synonyms
Conceptual and Syntactical Level

<xml version="1.0" encoding="UTF-8" />
<xmlns:hl7v3="http://www.w3.org/TR/WD-xlink" xmlns:hl7="urn:hl7-org:v3"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"/>

<xsiType="CE" code code="267629007"
codeSystem="2.16.840.1.113883.6.96"
codeSystemName="SNOMED CT"
codeSystemVersion="0601Core"
displayName="blurred vision NOS"
originalText="blurry vision">
</code>
HL V3 Class Object

References for each object

Frequency for each object

References for each object
Clinical Domain Object

<table>
<thead>
<tr>
<th>Name</th>
<th>Public Health Case Event</th>
<th>Incubation Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foodborne botulism investigation</td>
<td>Foodborne botulism</td>
<td>Botulism incubation period</td>
</tr>
</tbody>
</table>

**Investigated Agent**

- Botulism toxins

**Lab Observation**

- C. botulinum toxin detected in stool
- C. botulinum toxin detected in gastric aspirate
- C. botulinum toxin detected in food
- C. botulinum toxin detected in serum
- C. botulinum toxin E detected in stool
- C. botulinum toxin E detected in gastric aspirate
- C. botulinum toxin E detected in food
- C. botulinum toxin E detected in serum
- C. botulinum toxin A detected in stool
- C. botulinum toxin A detected in gastric aspirate
- C. botulinum toxin A detected in food
- C. botulinum toxin A detected in serum
- C. botulinum toxin F detected in stool

**Condition Observation**

- Diplopia
- Blurred vision
- Bulbar weakness
- Symmetric descending paralysis
- Dry mouth
- Dysphagia
- Dysarthria
- Muscle weakness

**Procedures**

- Tension test with electromyographic recording
- Diagnostic lumbar puncture

**Therapeutic Medication**

- BOTULISM ANTITOXIN TRIV A,B & E 5000 UNT Injectable, B & E

[Image]
<table>
<thead>
<tr>
<th>Name</th>
<th>ClassCode</th>
<th>RiskCode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Botulism toxins</td>
<td>MAT material</td>
<td>BHZ biohazard</td>
</tr>
</tbody>
</table>

HandlingCode: CREF Critical refrigerated temper...

DeterminerCode: INSTANCE specific

Causative Agent:
- Botulinum toxin type A
- Botulinum toxin type B
- Botulinum toxin type E
- Botulinum toxin type F

Citations:
- Control of Communicable Diseases
- http://www.dshs.state.tx.us/lab/mil
- CDC Public Health Image Library (PHIL)

Documentation:
Image from CDC/Dr. George Lombard

Image:
- c:\ontoreason_images\botox.jpg
Laboratory Observation HL7 V3 mapped to V2

- OBX-3
- OBX-8
- OBX-7
- OBX-5
- SPM-4

Text:
C botulinum toxin detected in stool

Documentation:

C botulinum toxin detected in stool

LabObservation.code:
C bot Tox XXX Qi

InterpretationCode:
A Abnormal

MethodCode:
0025 Bioassay

TargetSiteCode:
Stool specimen

Observation.Value:
Detected

InterpretationStatus:
Not detected

ClassCode:
observation

MoodCode:
event (occurrence)
Map HL7 Message segments to Ontology Slots

<table>
<thead>
<tr>
<th>LabObservation.code</th>
<th>This slot is populated by an instance of the LOINC coded value that represents the ordered test.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value Type</td>
<td>Class</td>
</tr>
<tr>
<td>Allowed Superclasses</td>
<td>LOINC Code System</td>
</tr>
<tr>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>Inverse Slot</td>
<td></td>
</tr>
</tbody>
</table>

- <OBX.2>NM</OBX.2>
  - <OBX.3>
    - <CE.1 />
      - <CE.2>CSF protein conc</CE.2>
      - <CE.5>CSF protein mg%</CE.5>
  </OBX.3>
- <OBX.4>34</OBX.4>
- <OBX.5>
  - <NM.1>34</NM.1>
  </OBX.5>
- <OBX.6>
  - <CE.1>mg/dl</CE.1>
  </OBX.6>

:DIRECT_TYPE
     :STANDARD-SLOT
     :HL7V2_EQ

:SL ot-CONSTRAINTS

- <OBX.4> CRM
- <OBX.3> CRM
- <OBX.4> CRM

Vocabulary Recommended
- Laboratory Test Name

Documentation
This slot is populated by an instance of the LOINC coded value that represents the ordered test.

Translation/Requirement
None
Ontology Services Platform
Technical Foundations

Platform Models
- Enterprise PHIN SOA
- Web Services
- Application Libraries
- LexPHIN Database

Application Models
- Individual Reasoners Patterns - Languages
- Intelligence & Analytics Workbench - Tools
- CTS & LexPHIN Services - Standards

Domain Models
- PH Reference Ontology
- PHIN VS
- BioSense Msg HL7 V2.x

Message Structure
Ontology Extraction

- Creating an application ontology from the reference ontology
  - Identify the core ontology classes
  - Create an object representation that maintains the ontology data
  - Generate cross reference indexes for core relationships
    - Lab tests to case investigations
    - Organism/Agent to case investigations
    - Other significant relationships
  - Identify “Used” vocabulary
    - Create vocabulary subsets that identify specific vocabularies concepts that are used within the ontology
    - Create code to code mapping indexes
- This produces a general purpose extraction that is suitable for various purposes
Ontology Extraction

- Additional activities performed for specific problem solutions
  - Inclusion of additional vocabulary value sets
  - Generation of additional vocabulary indexes to maintain certain parent/child relationships
  - Incorporation of certain additional term mappings
    - Alternate spellings
    - Concept mappings to syndrome/sub-syndrome
  - Generation of text search algorithms
- Loadable data married with functional API
  - Java object serialized for easy loading
  - Java API providing lookup/query functionality
Ontology Representation

- **Jess rule engine representation**
  - The Jess rule engine utilizes an enhanced RETE algorithm to provide an execution platform for declarative rule base
  - Data in Jess is represented as a set of declared facts
    - Facts can be either structured or unstructured
    - Ontology data is represented as a set of instance data
  - The ontology can either be expressed as a script or loaded directly into the rule engine at runtime

- **Rule definition**
  - Rules which describe core case definitions are constructed
  - The ontology facts are merged with the core set of rules to provide the base representation for the entire ontology
Message Analytics Workbench
Model Driven Expert System

- Public Health Reference Ontology
  - Public health domain model
  - Highly constrained standardized vocabulary

- Institutional Knowledge
  - Clinical reference material
  - Expert knowledge representation

- Calculated Knowledge
  - Statistical information
  - Empirical evidence

- Web-Service based application components

- Reasoning Patterns
- Platform Descriptions
BioSense Message Data Source

- HL7 Version 2.5
- XML representation
- Laboratory (ORU) message
- Spinal fluid protein

```xml
<?xml version="1.0" encoding="ISO-8859-1" ?>

<ORU_R01>
  + OBS
    - ORU_R01.OBSERVATION
      + OBS
        + OBS.1
          + OBR.1
            + OBR.2
              + OBR.3
                + OBR.4
                  + OBR.5
                    + OBR.6
                      + OBR.7
                        + OBR.8
                          + OBR.9
                            + OBR.10
                              + OBR.11
                                + OBR.12
                                  + OBR.13
                                    + OBR.14
                                      + OBR.15
                                        + OBR.16
                                          + OBR.17
                                            + OBR.18
                                              + OBR.19
                                                + OBR.20
                                                  + OBR.21
                                                    + OBR.22
                                                      + OBR.23
                                                        + OBR.24
                                                          + OBR.25
                                                            + OBR.26
                                                              + OBR.27
                                                                + OBR.28
                                                                  + OBR.29
                                                                    + OBR.30
                                                                      + OBR.31
                                                                        + OBR.32
                                                                          + OBR.33
                                                                            + OBR.34
                                                                              + OBR.35
                                                                                + OBR.36
                                                                                  + OBR.37
                                                                                      + OBR.38
                                                                                       + OBR.39
                                                                                        + OBR.40
                                                                                         + OBR.41
                                                                                          + OBR.42
                                                                                           + OBR.43
                                                                                            + OBR.44
                                                                                             + OBR.45
                                                                                              + OBR.46
                                                                                               + OBR.47
                                                                                               </OBR>
                                                                                           </OBR.15>
                                                                                          </OBR.22>
                                                                                           </OBR.23>
                                                                                           </OBR.24>
                                                                                       </OBR.25>
                                                                                      </OBR.26>
                                                               </OBR.27>
                                                              </OBR.28>
                                                             </OBR.29>
                                                            </OBR.30>
                                                           </OBR.31>
                                                          </OBR.32>
                                                         </OBR.33>
                                                        </OBR.34>
                                                       </OBR.35>
                                                      </OBR.36>
                                                     </OBR.37>
                                                    </OBR.38>
                                                   </OBR.39>
                                                  </OBR.40>
                                                 </OBR.41>
                                                </OBR.42>
                                               </OBR.43>
                                              </OBR.44>
                                             </OBR.45>
                                            </OBR.46>
                                         </OBR.47>
                                        </OBR.48>
                                      </OBR.49>
                                    </OBR.50>
                                  </OBR.51>
                                </OBR.52>
                              </OBR.53>
                            </OBR.54>
                          </OBR.55>
                        </OBR.56>
                      </OBR.57>
                    </OBR.58>
                  </OBR.59>
                </OBR.60>
              </OBR.61>
            </OBR.62>
          </OBR.63>
        </OBR.64>
      </OBR.65>
    </OBS>
  </ORU_R01>
</ORU>
```
Demonstration of Basic Platform

- Message Processing
- User Profile Data Entry
- Intelligence & Analytics Dashboard
- Dynamic Syndrome Definition
- Knowledge Bus
  - Message Classification
  - Patient Msg. Correlation/Classification
  - Cross Patient Correlation
  - Cross Patient Correlation
Message Analytical Workbench

HL/7 Message Source

HL7 Message Parser  Fact Generator  OTR Ontology

Analytics Interface

Knowledge Board

Dashboard

Reasoner Controller

Situation Awareness (multiple cases)  Message Classification Reasoner  Patient Correlation Reasoner

Classification Definitions  Routing instructions
Anthrax Investigation (Inhalation)

Citations

- Anthrax as a biological weapon JAMA 2002;287:2236-2252
- www.aafp.org/afp/20030501/1927.html

HL7 Representation

public health case [CASE] event (occurrence) [EVT]
Reasoner results
How the Rules Work

- JESS template is like a class in Java
- Template can but does not have to declare attribute type, default values, and if an attribute is a single value or a list
- Facts asserted into the expert system directly from the ontology - based upon the template
Interaction of Ontology And Rules

• Information from the ontology and generated template facts imported into the expert system and operated on by a variety of reasoners.

• This way the reasoner knowledgebase can have a relatively small footprint vs the reference ontology.
(defrule ClinicalFindingConditionMatchesInitial
  (Classifier-State DevelopFindingsAndEndorsements)
  (observation (obsId ?obsId)(msgId ?mId)(patientId ?patientId)(chiefComplaint ?chief)
    (code ?observation) (obsType ClinicalFinding) (dataQualityFactor ?quality)
    (dqfReason ?qualityReason) (originationDate ?oDate) (originationDateType ?oType))
  (nnd-condition (cond-code ?condCode) (description ?desc))
  (not (BSFinding (msgId ?mId)(finding ?condCode)))
=>
  (assert (BSFinding (msgId ?mId)(patientId ?patientId) (findingId *(ClassId*))
    (finding ?condCode) (findingDesc ?desc)
    (originationDate ?oDate)(originationDateType ?oType))

  (assert (Endorsement (msgId ?mId)(findingId *(ClassId*) (findingCorrelation ?ratio)
    (findingType Condition)(finding ?condCode)(findingProb ?prob)
    (endorsementId (+ *(ClassId* 1))(endorsement
    ?*Supportive*)(endSymbol *Support*)
    (endorsementType *ClinicalFinding*)(rule
    *ConditionMatch*)(endorsementContext ?context )
    (obsId ?obsId)(obsCode ?observation)(obsQuality
    ?quality)(explaination ?qualityReason )))
  (bind *(ClassId* (+ 2 *(ClassId*)))))
A couple of things to remember

- This is an Intelligence and Analytics toolkit
  - Used to exploit the expert knowledge of the organization to provide simple to configure application components
  - Real-time processing
  - Historical data for analysis, knowledge discovery and re-classification
  - Findings can be reused to tune and validate real-time processing
- Classification tools are based upon a very quick assessment generalized across all conditions
- The classification weights can be greatly improved based upon empirical data analysis
- Algorithms are simple to tune and extend (including geo-spatial and temporal services)
- The use-cases were made from some limited set of assumptions
- We used a condition centric analysis
Next Steps

- Add additional domain centric rules for better agent classification
- Overcome limitations of ontology size and maintenance issues by subdividing into smaller ontologies
- Apply a novel technique to use the best aspects of Frames and OWL structures
  - (see the demo)
- Develop simple domain expert editing tools for rules and knowledge
Questions and Answers