



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

Cecil Lynch<sup>1,2</sup>, Craig Cunningham<sup>1</sup>, Eric  
Schripsema<sup>1</sup>, Tim Morris<sup>3</sup>, Barry Rhodes<sup>3</sup>

# 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

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

# 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

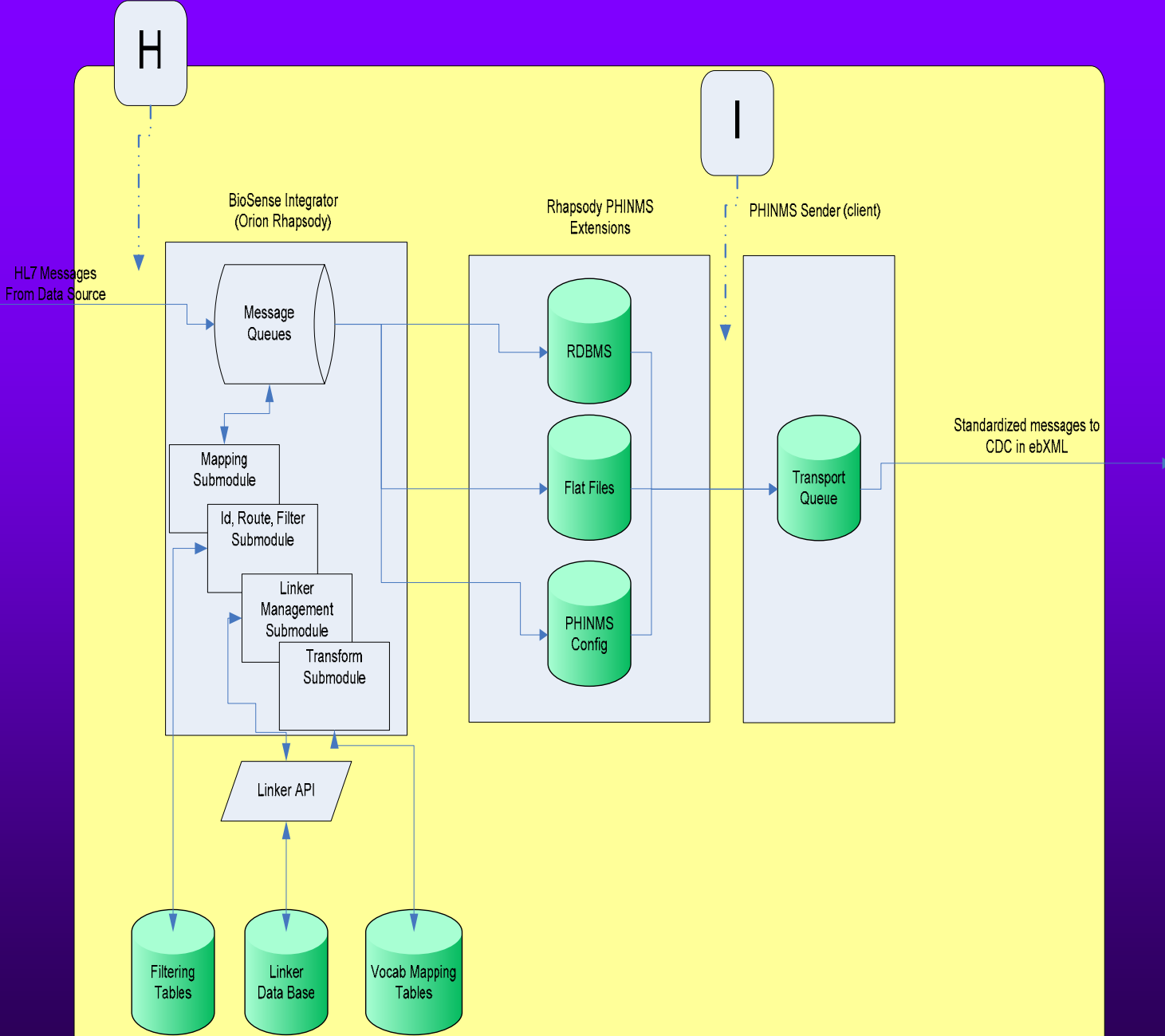
- Data mapped to 11 syndrome categories
  - Botulism-like
  - Fever
  - Gastrointestinal
  - Hemorrhagic illness
  - Localized cutaneous lesion
  - Lymphadenitis
  - Neurological
  - Rash
  - Respiratory
  - Severe illness/death
  - Specific infection
- 79 sub-syndrome categories

# Watch what you ask for!

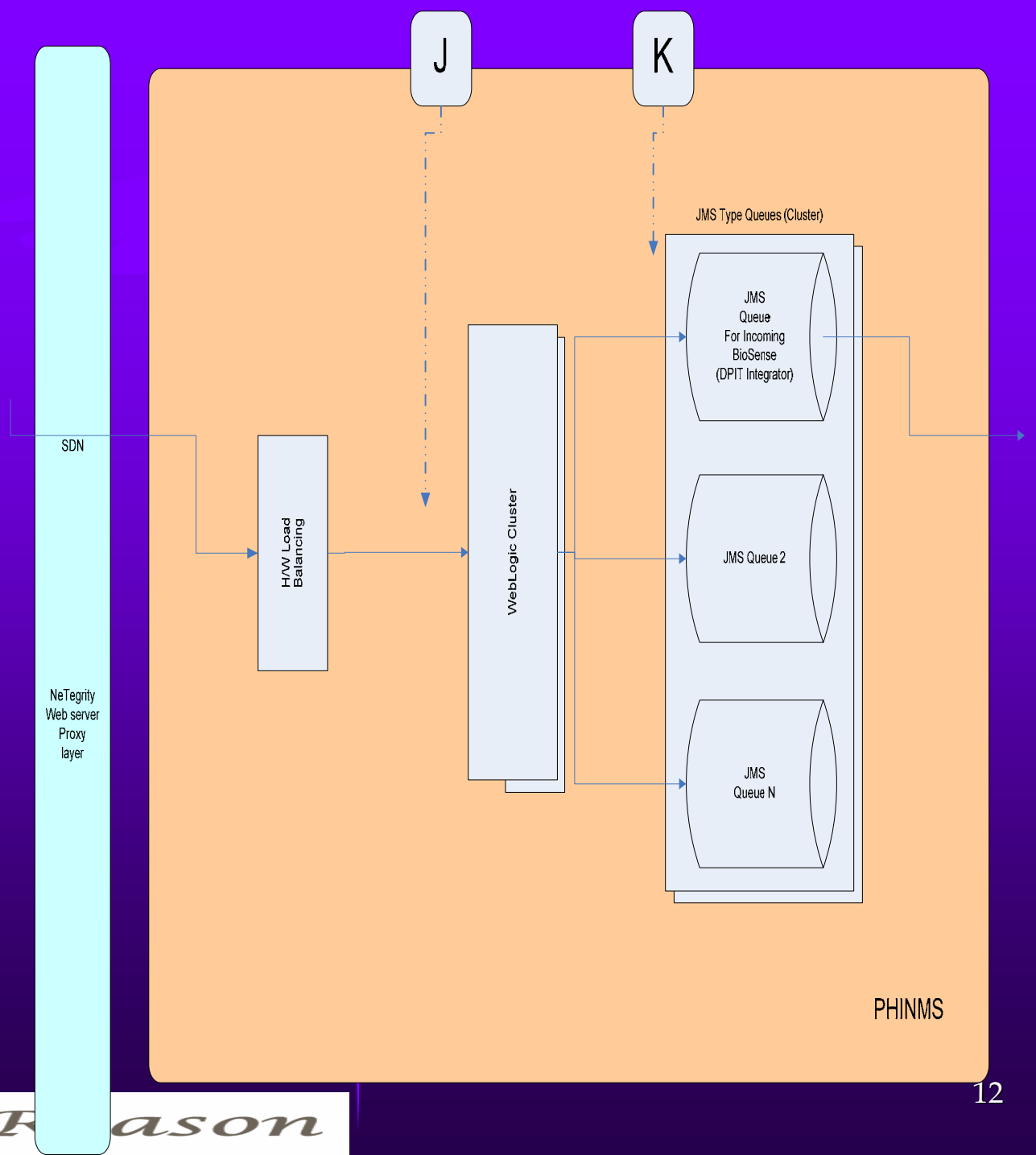
- BioSense message volume capacity today
  - 837 messages a second
  - >72 million messages a day
- How does an epidemiologist review that volume of data?
- How do you link messages to an individual over time to refine the diagnostic info?

# Current BioSense Framework

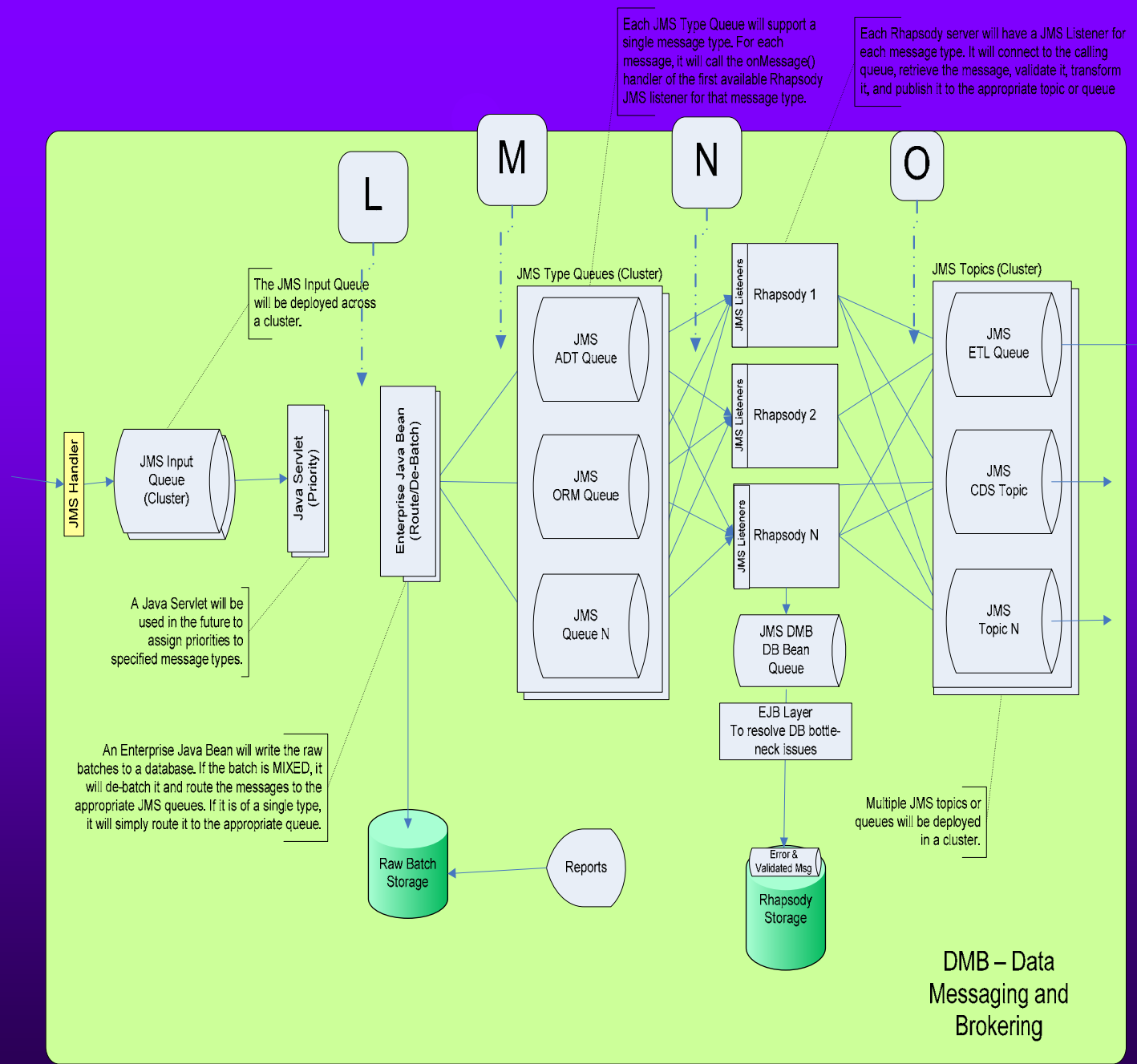
# Message Processing



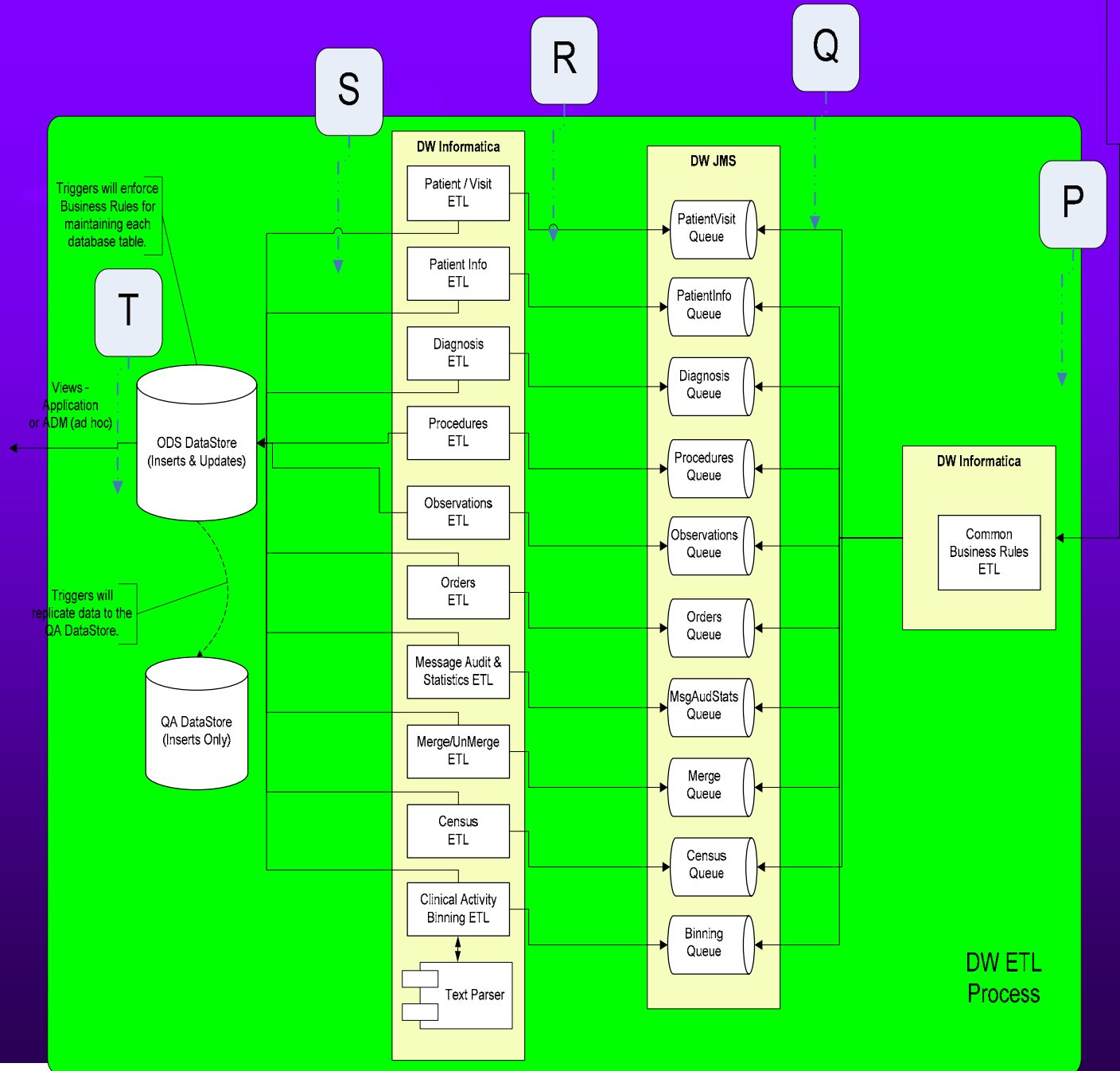
# Load Balancing



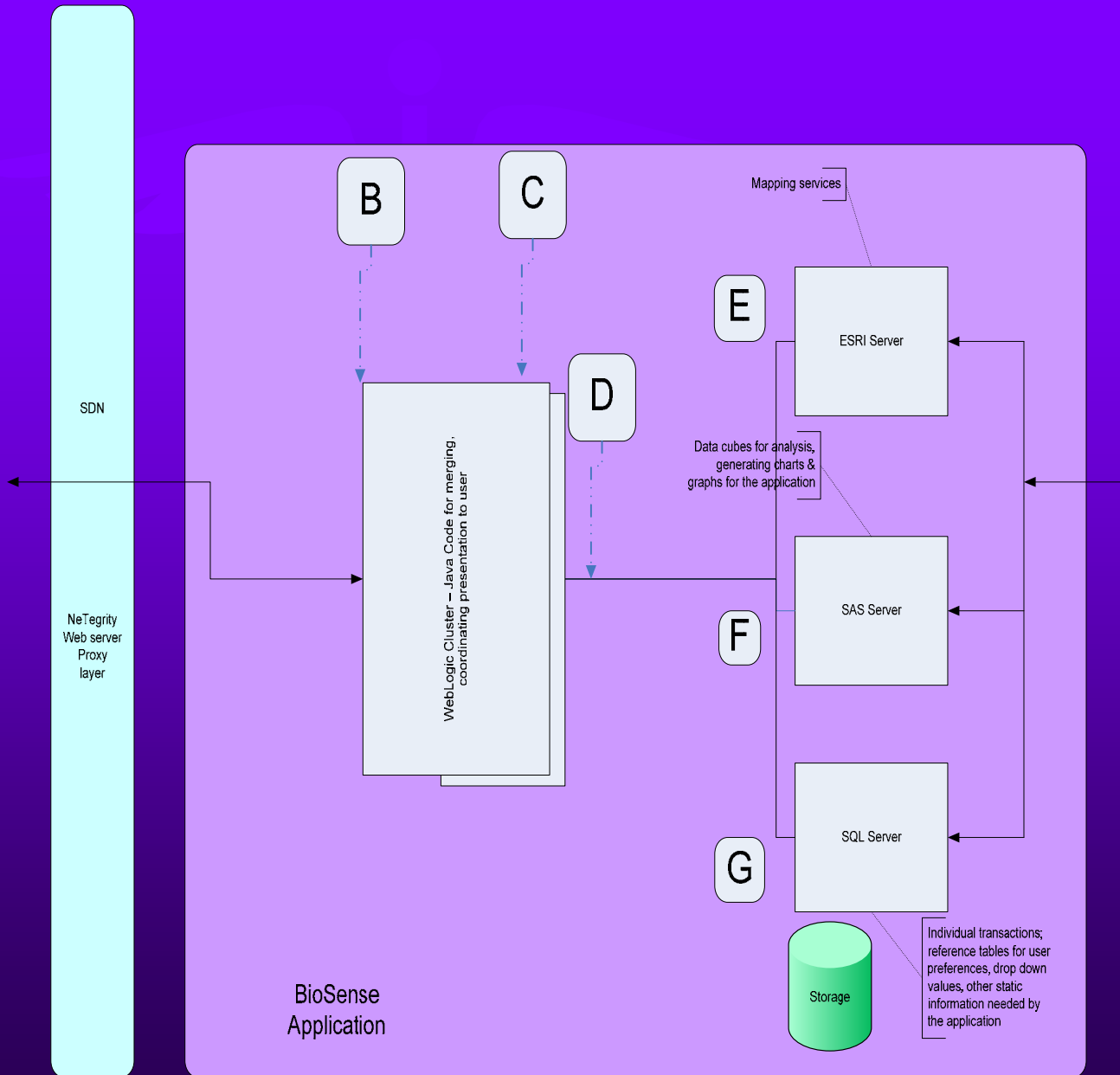
# Message Type Filter



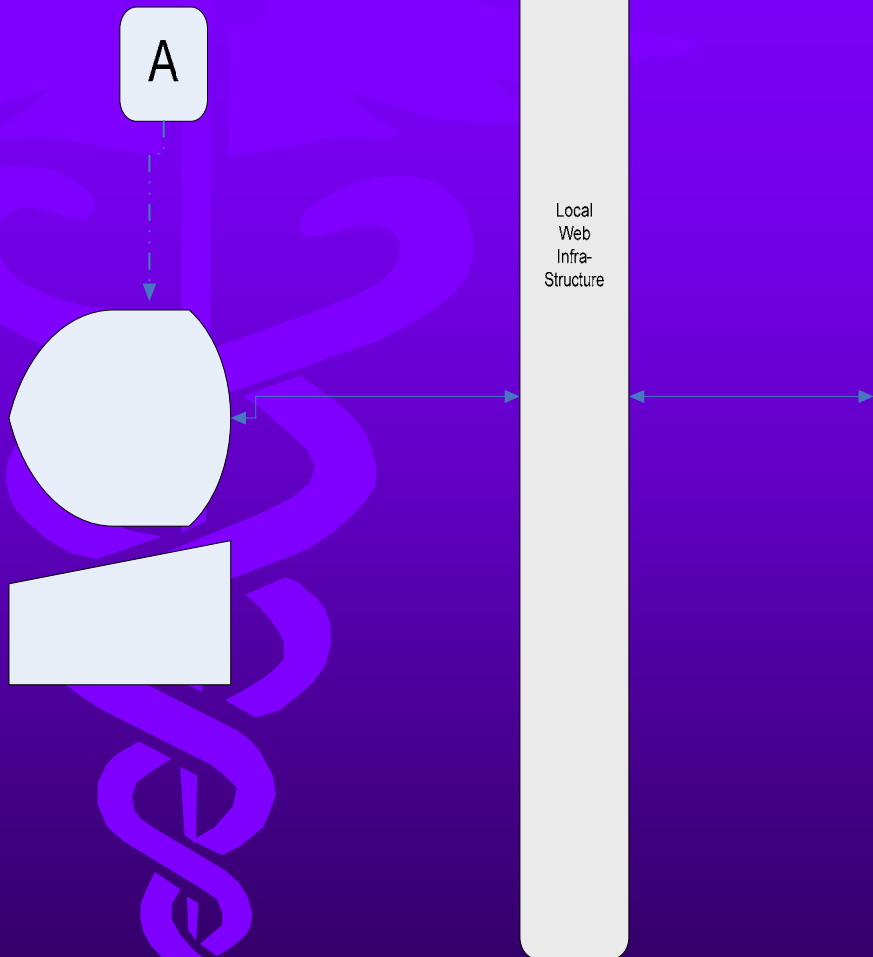
# ETL Processing



# AV and OTP



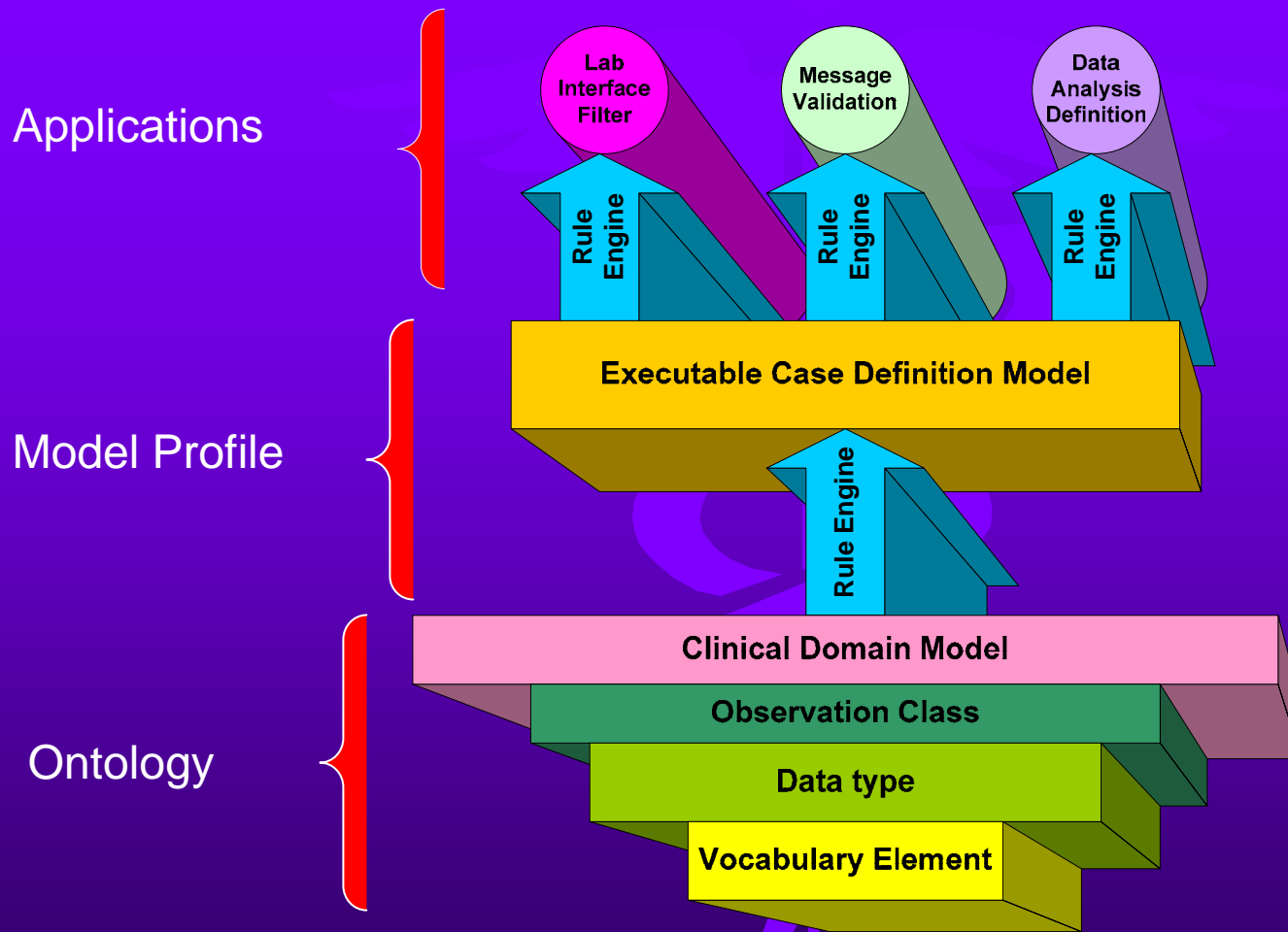
# End User Views

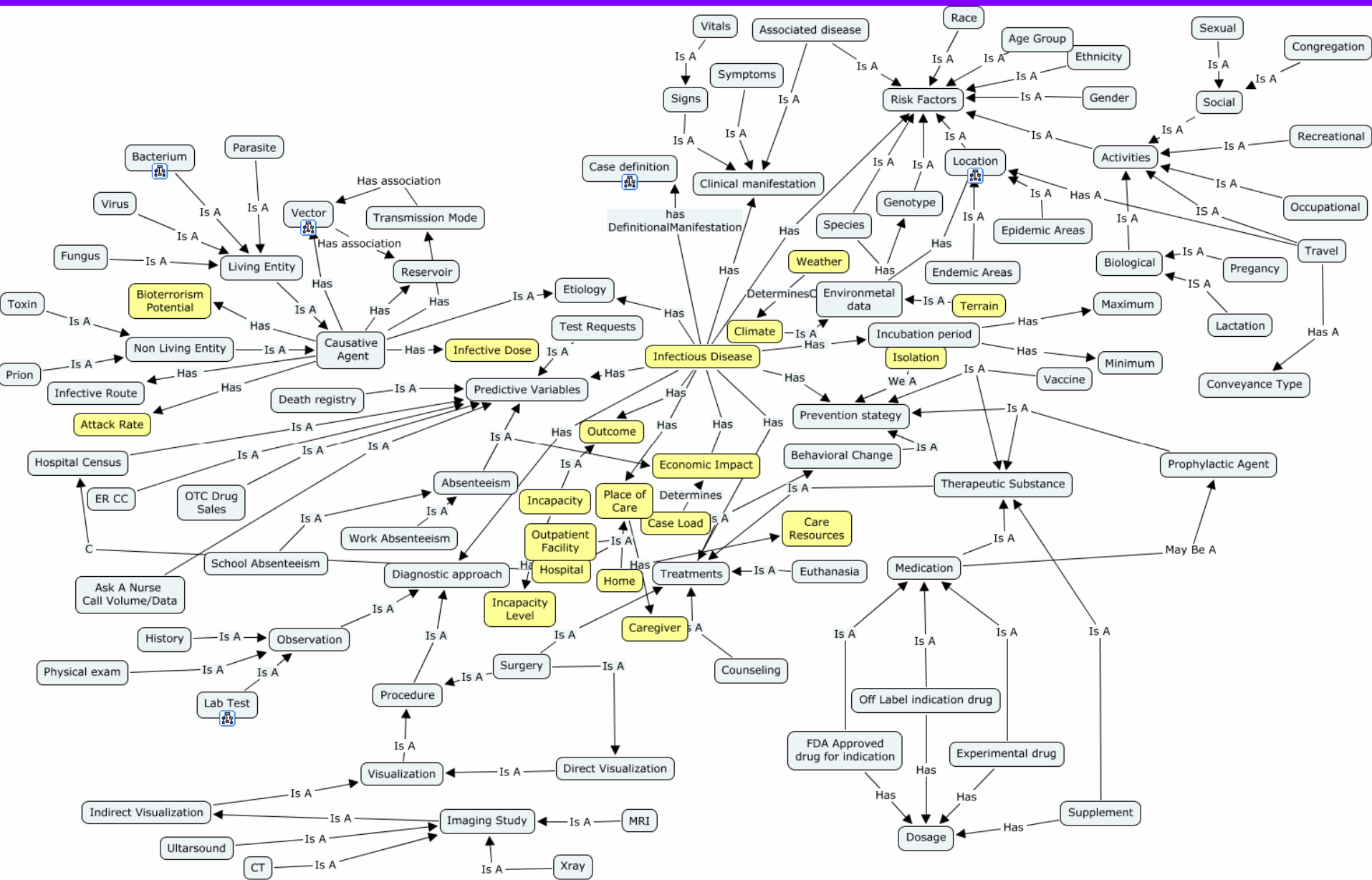


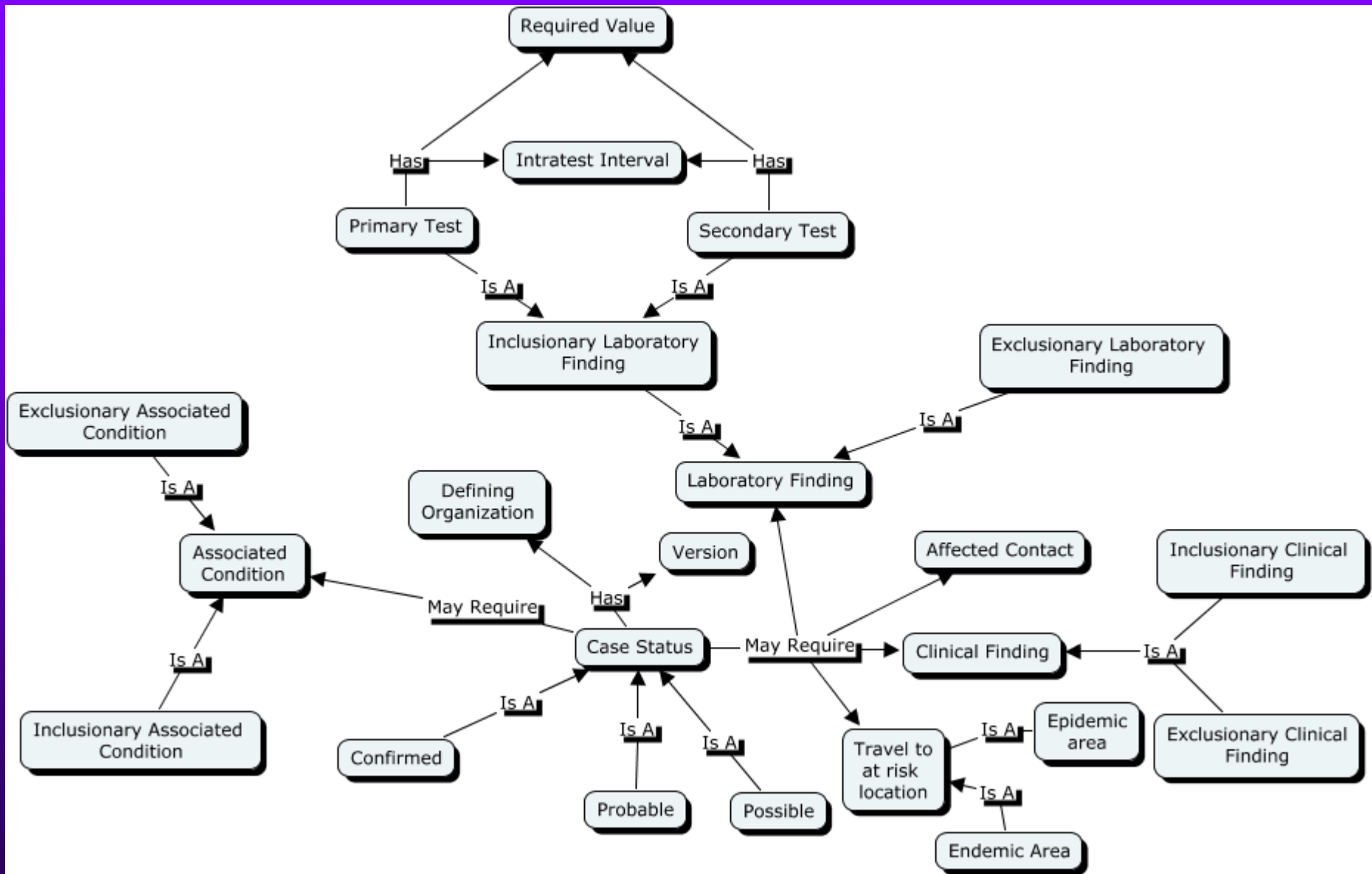
BioSense Users  
On  
The World Wide Web



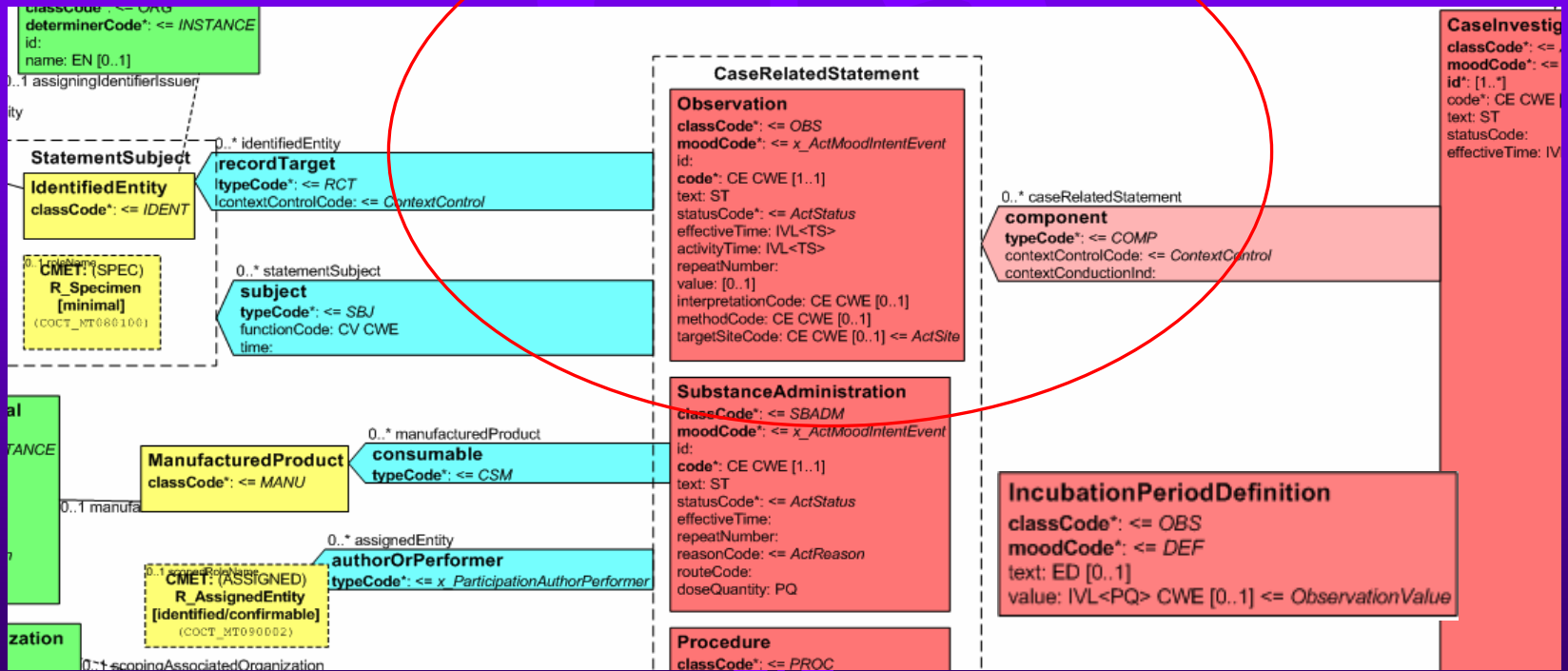
# The OntoReason PH Ontology







# Information Model



# Concept In HL7 V3 DataType

The screenshot displays the OntoReason interface for a concept named 'Cough'. The interface is organized into several panels:

- Code:** 49727002
- CodeSystem:** 2.16.840.1.113883.6.96
- CodeSystemName:** SNOMED CT
- CodeSystemVersion:** 0601core
- 01/31/2006**
- Qualifier:**
- ConceptStatus:** Active
- DisplayName:** Cough
- Parent:** Functional finding of respiratory tract
- OriginalText:** Cough
- Superclasses:** Functional finding of respiratory tract
- Subclasses:** 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
- Translation:** Cough (finding), Observation of cough, Finding of cough, Cough
- Name:** Lynch82824016SN
- BioSenseMatch:** toss 1221, couhg 300, 7862 3902, Cough 3298

Callouts in the image identify the following elements:

- Code:** Points to the Code field (49727002).
- Term:** Points to the DisplayName field (Cough).
- Parent:** Points to the Superclasses field (Functional finding of respiratory tract).
- Children:** Points to the Subclasses list.
- BioSense Terms:** Points to the BioSenseMatch list.
- Other code Systems and synonyms:** Points to the Translation list.

# Conceptual and Syntactical

## Level

Condition Observation

- Diplopia
- Blurred vision
- Blurred vision (instance)

Text

Blurred vision

ClinicalObservation.1

- Blurred vision NOS

Low 90.0 Width

● Blurred vision NOS (instance of Disorders\_MetaClass, internal name is Lynch3.)

Code	CodeSystem	CodeSystemName
267629007	2.16.840.1.113883.6.96	SNOMED CT
CodeSystemVersion	DateCreated	Qualifier
0601core	01/31/2006	
ConceptStatus	DisplayName	
Active	Blurred vision NOS	
DateChanged	OriginalText	
	Blurred vision NOS	

```
<?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>
```



# Clinical Domain Object

The screenshot displays a clinical domain object interface with the following sections:

- Name:** Foodborne botulism investigation
- Public Health Case Event:** Foodborne botulism
- Incubation Period:** Botulism incubation period
- Investigated Agent:** Botulism toxins (highlighted with a red callout box labeled "Nested MetaClass")
- LabObservation:**
  - 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
- Vaccine:** (Empty)
- Radiology Observation:** CT of head normal
- Condition Observation:**
  - Diplopia
  - Blurred vision
  - Bulbar weakness
  - Symmetric descending paralysis
  - Dry mouth
  - Dysphagia
  - Dysarthria
  - Muscle weakness
- Procedures:**
  - Tensilon test with electromyographic recording
  - Diagnostic lumbar puncture
- Therapeutic Medication:** {BOTULISM ANTITOXIN TRIV A,B&E 5000 UNT Injectable, E

Investigated Agent 🔍 ✨ ⬇️ ⬆️

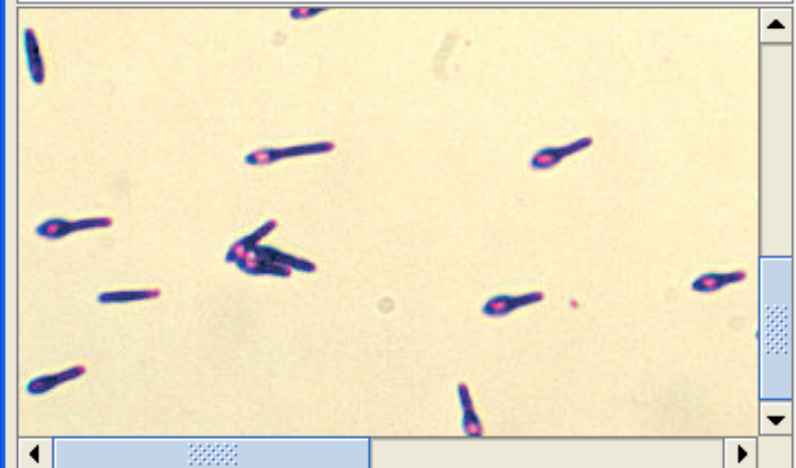
● Botulism toxins

● Botulism toxins (instance of EntityClassMaterial\_MetaClass, internal name is 0... [min] [max] [close])

<b>Name</b>	<b>ClassCode</b> 🔍 ✨ ⬇️ ⬆️	<b>RiskCode</b> 🔍 ✨ ⬇️ ⬆️
Botulism toxins	● MAT material	● BHZ biohazard
<b>HandlingCode</b> 🔍 ✨ ⬇️ ⬆️	<b>DeterminerCode</b> 🔍 ✨ ⬇️ ⬆️	<b>Causative Agent 246</b> 🔍 ✨ ⬇️ ⬆️
● CREF Critical refrigerated temper...	● INSTANCE specific	● Botulinum toxin type A ● Botulinum toxin type B ● Botulinum toxin type E ● Botulinum toxin type F
<b>Citations</b> 🔍 ✨ ⬇️ ⬆️	<b>Documentation</b>	
● Control of Communicable Diseases ● <a href="http://www.dshs.state.tx.us/lab/mr">http://www.dshs.state.tx.us/lab/mr</a> ● CDC Public Health Image Library (Ph	Image from CDC/Dr. George Lombard	

**Image**

c:\ontoreason\_images\bottox.jpg



**NegationValue** 🔍 ✨ ⬇️ ⬆️

# Laboratory Observation HL7 V3 mapped to V2

C botulinum toxin detected in stool (instance of LabObservation\_MetaClass, in...)

Text: C botulinum toxin detected in stool

Documentation:

Citations:  
● <http://www.cdc.gov/epo/dphsi/cas>  
● <http://www.dshs.state.tx.us/lab/mr>

LabObservation.cod: ● C bot Tox XXX QI (OBX-3)

InterpretationCode: ● A Abnormal

MethodCode: ● 0025 Bioassay (OBX-17)

ClassCode: ● observation

MoodCode: ● event (occurrence)

Observation.Val: ● Detected (OBX-5)

InterpretationRange: ● Not detected (OBX-7)

TargetSiteCode: ● Stool specimen (OBX-8)

SPM-4

# Map HL7 Message segments to Ontology Slots

```
<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>
```

LabObservation.code

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

Value Type  
Class

Allowed Superclasses  
● LOINC Code System

Cardinality  
 required at least   
 multiple at most

Minimum  Maximum  Inverse Slot

:DIRECT-TYPE  
● :STANDARD-SLOT  
● HL7V2\_EQ

:SLOT-CONSTRAINTS

as HL7V2\_EQ

Name  
LabObservation.code

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

TranslationRequirement  
None

HL7V2 Field  
<OBR.4> ORM  
<OBX.3> ORU  
<OBR.4> ORU

Vocabulary Recommended  
● Laboratory Test Name

Default Values

Domain  
● LabObservation\_MetaClass



# 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

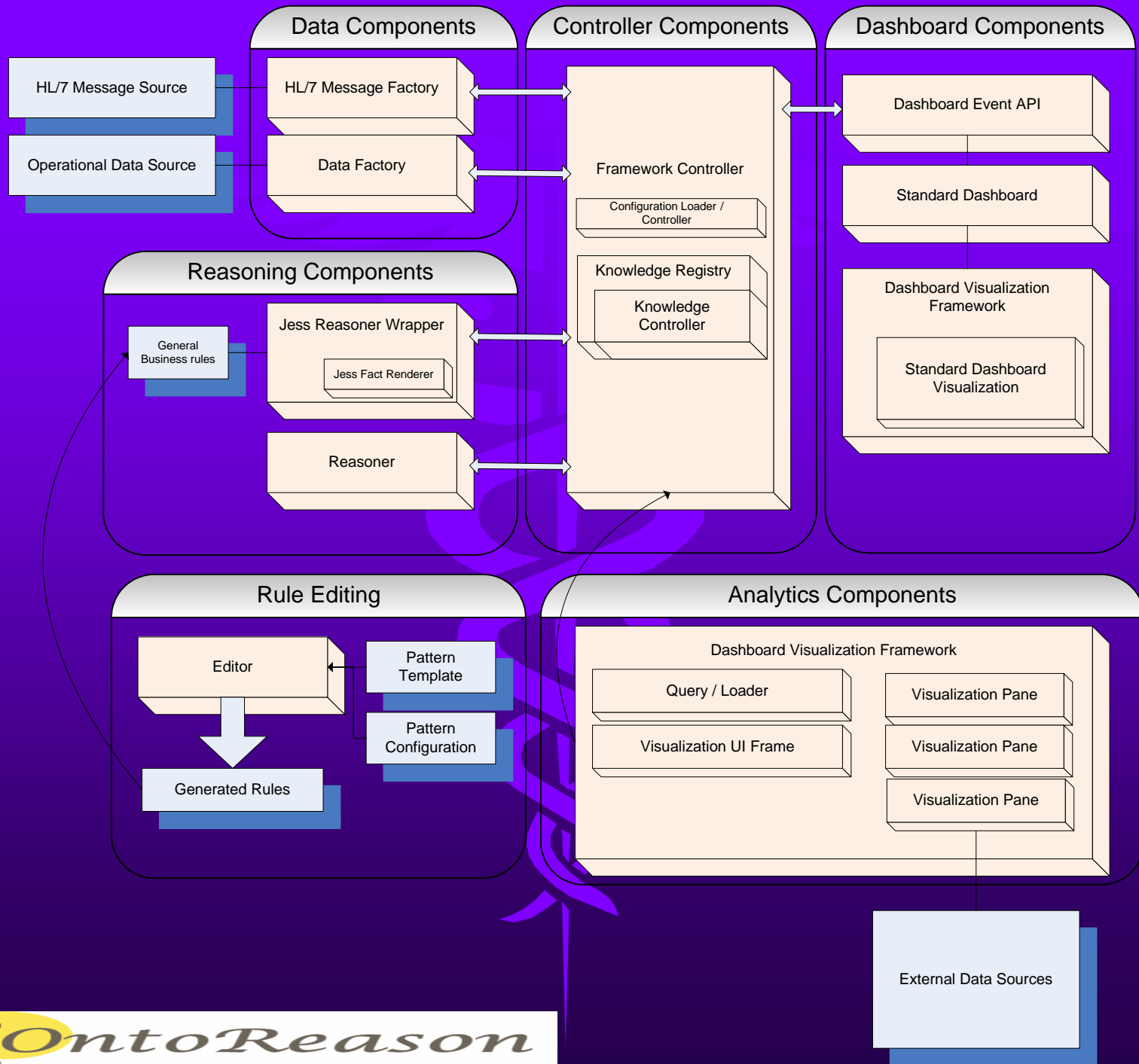
- **Creating an application ontology from the reference ontology**
  - Identify the core ontology classes
  - Create an object representation that maintains the ontology data
  - Generate cross reference indexes for core relationships
    - Lab tests to case investigations
    - Organism/ Agent to case 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

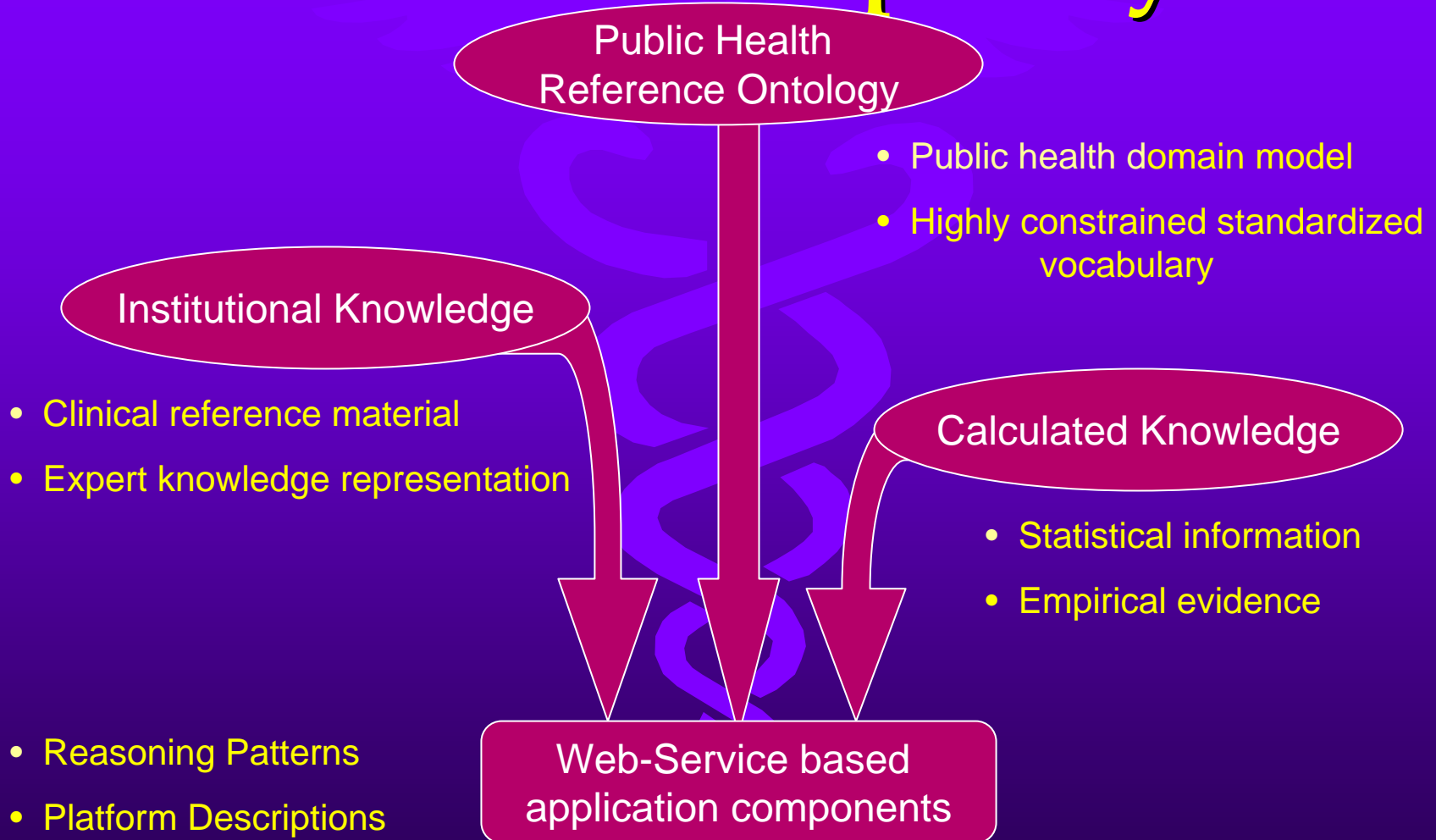
- **Jess rule engine representation**
  - The Jess rule engine utilizes an enhanced RETE algorithm to provide an execution platform for declarative rule base
  - Data in Jess is represented as a set of declared facts
    - Facts can be either structured or unstructured
    - Ontology data is represented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

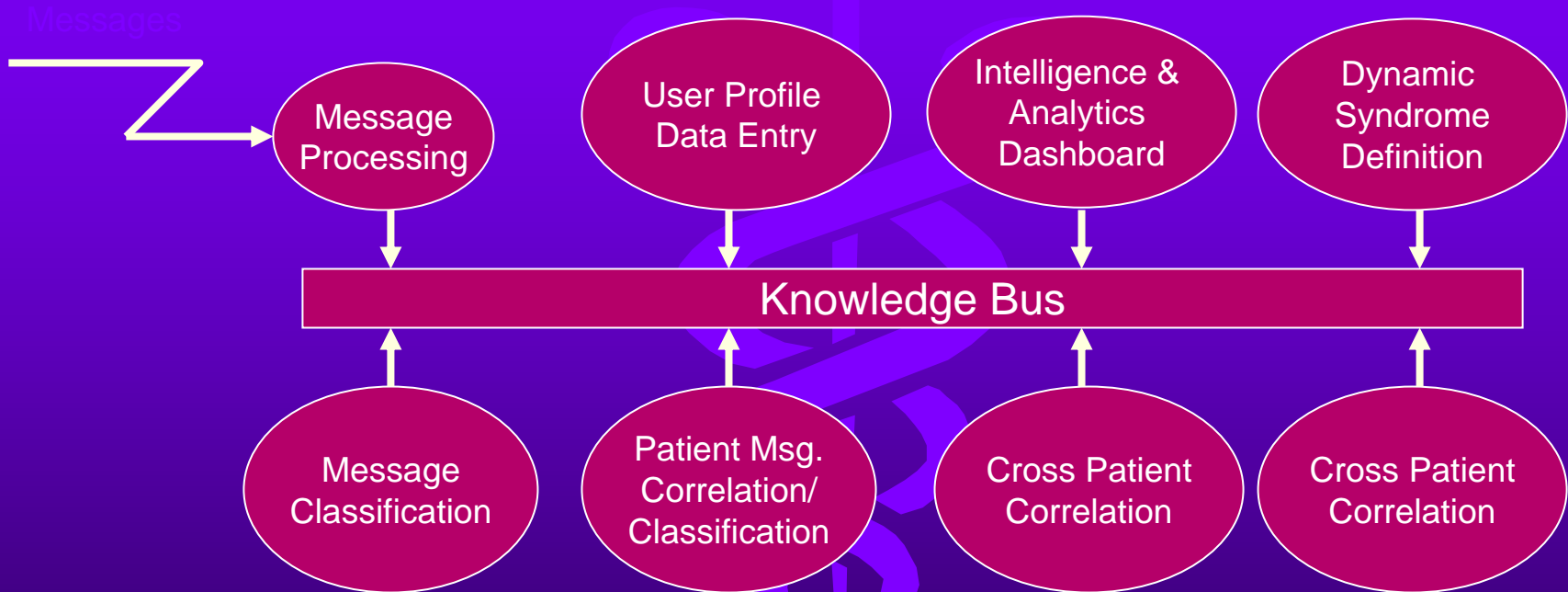


# BioSense Message Data Source

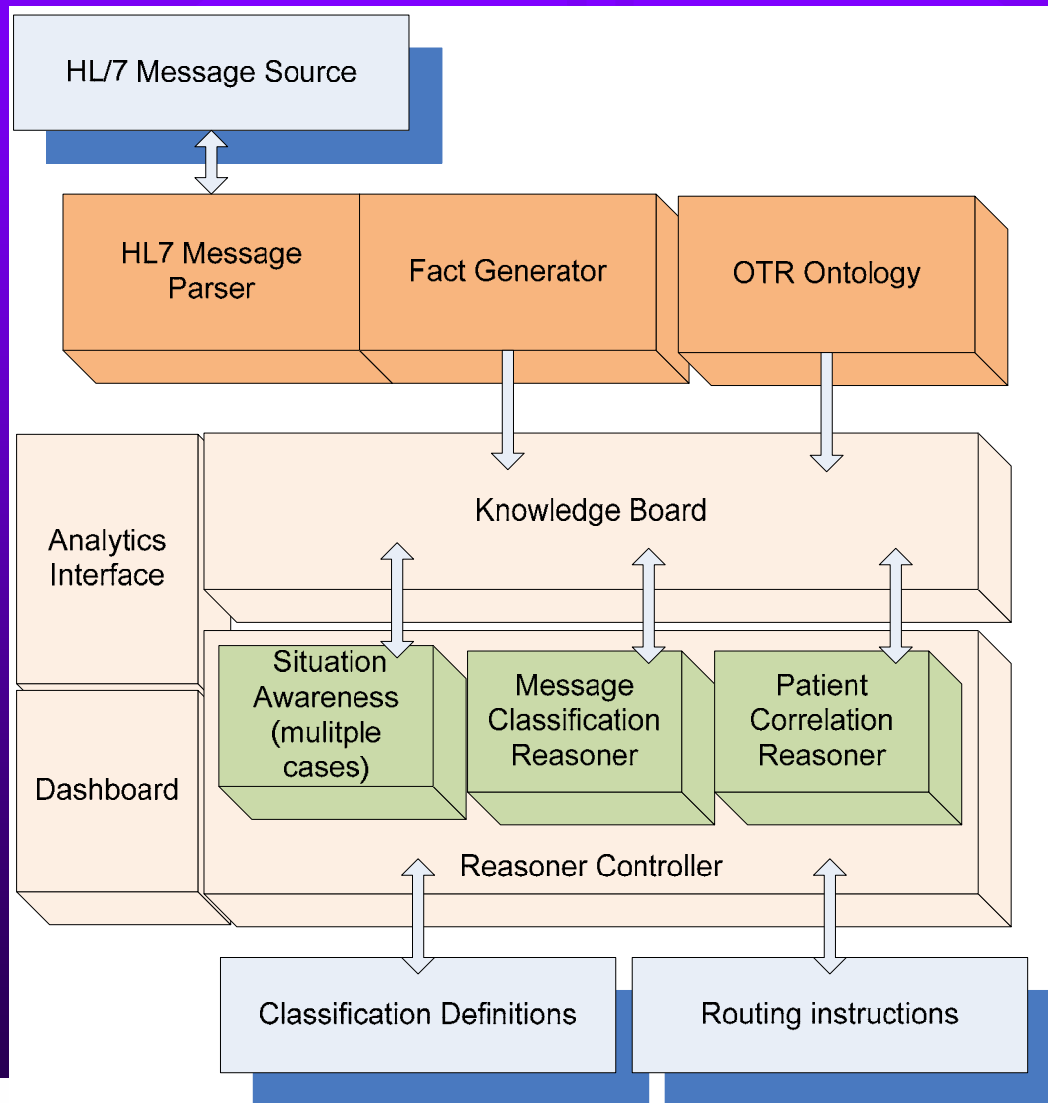
- HL7 Version 2.5
- XML representation
- Laboratory (ORU) message
- Spinal fluid protein

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- edited with XMLSpy v2007 sp2 (http://www.altova.com) by Cec
- <ORU_R01>
+ <MSH>
- <ORU_R01.PATIENT_RESULT>
+ <ORU_R01.PATIENT>
- <ORU_R01.ORDER_OBSERVATION>
- <OBR>
  <OBR.1>1</OBR.1>
  + <OBR.2>
  + <OBR.3>
  - <OBR.4>
    <CE.4>CSF protein</CE.4>
    <CE.5>CSF protein</CE.5>
  </OBR.4>
  + <OBR.7>
  + <OBR.8>
  + <OBR.14>
  - <OBR.15>
    - <SPS.1>
      <CWE.4>Spinal Fluid</CWE.4>
    </SPS.1>
  </OBR.15>
  + <OBR.22>
    <OBR.24>LA</OBR.24>
    <OBR.25>F</OBR.25>
  </OBR>
- <ORU_R01.OBSERVATION>
- <OBX>
  <OBX.1>1</OBX.1>
  <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>
```

# Demonstration of Basic Platform



# Message Analytical Workbench



File Edit Window Help

NND Data **DX Diag** **Cond Def** **Case Def** **Epi Sur** **Ter Match** **OntoReason**

**NND Selector**

Select: Anthrax investigation (inhalation)  Qualified?

- Inhalational anthrax 11389007
    - Transmission Modes
      - Airborne transmission (416380006)**
    - Incubation Periods
      - Anthrax incubation period (inhalation route)**
        - Period (Range) 1.0-60.0(day)
        - Period (Median) 1.0-7.0(day)
    - Clinical Findings
    - Laboratory Observations
    - Organisms/Agents
      - Bacillus anthracis (21927003)**
    - Radiology Observations

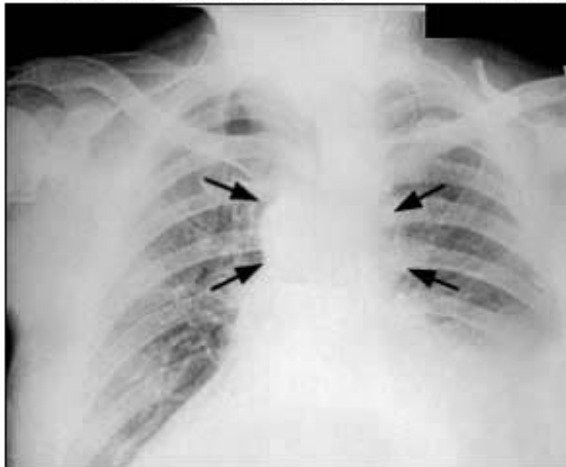
**Anthrax investigation (inhalation)**

**Citations**

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

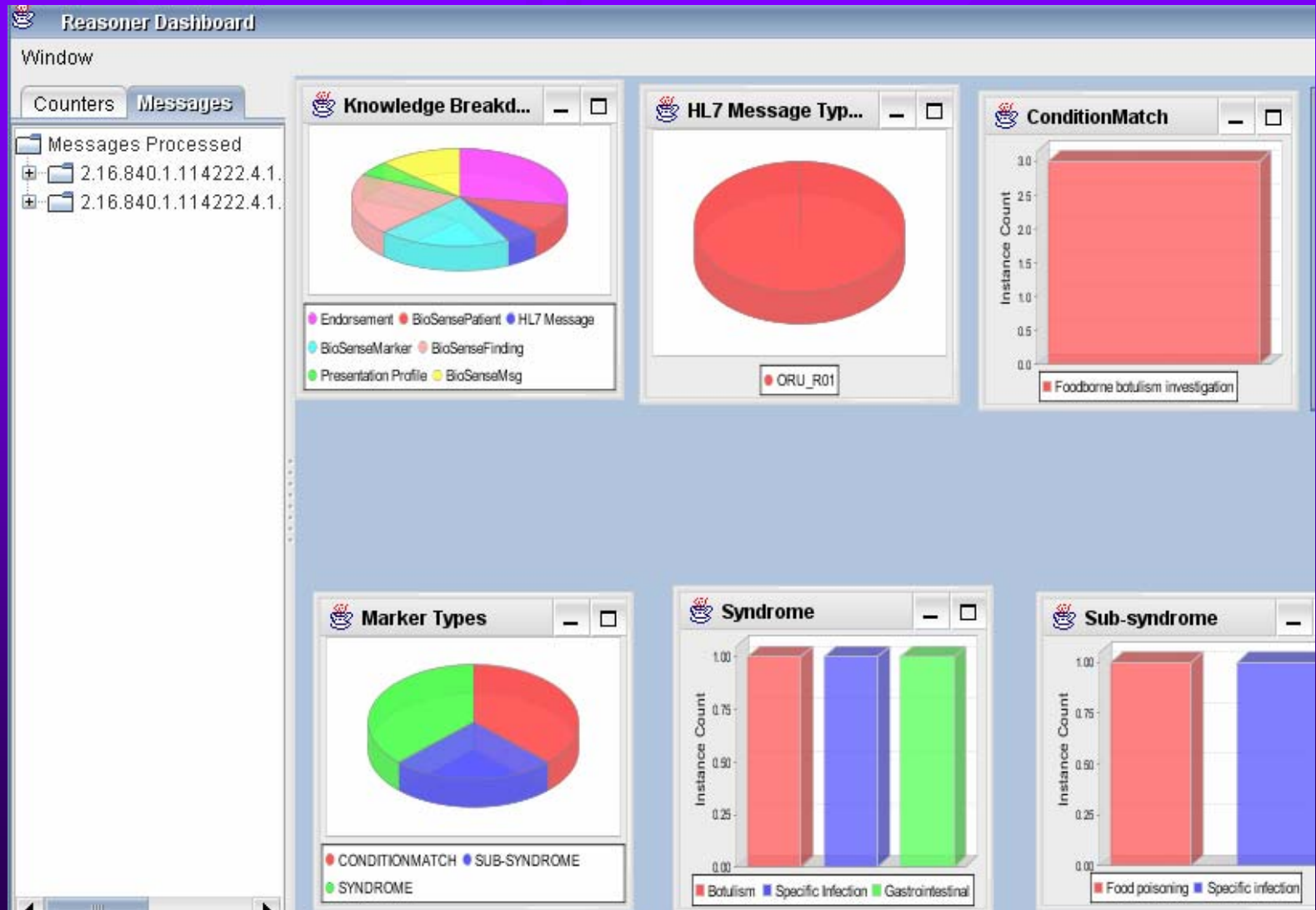
**HL7Representation**

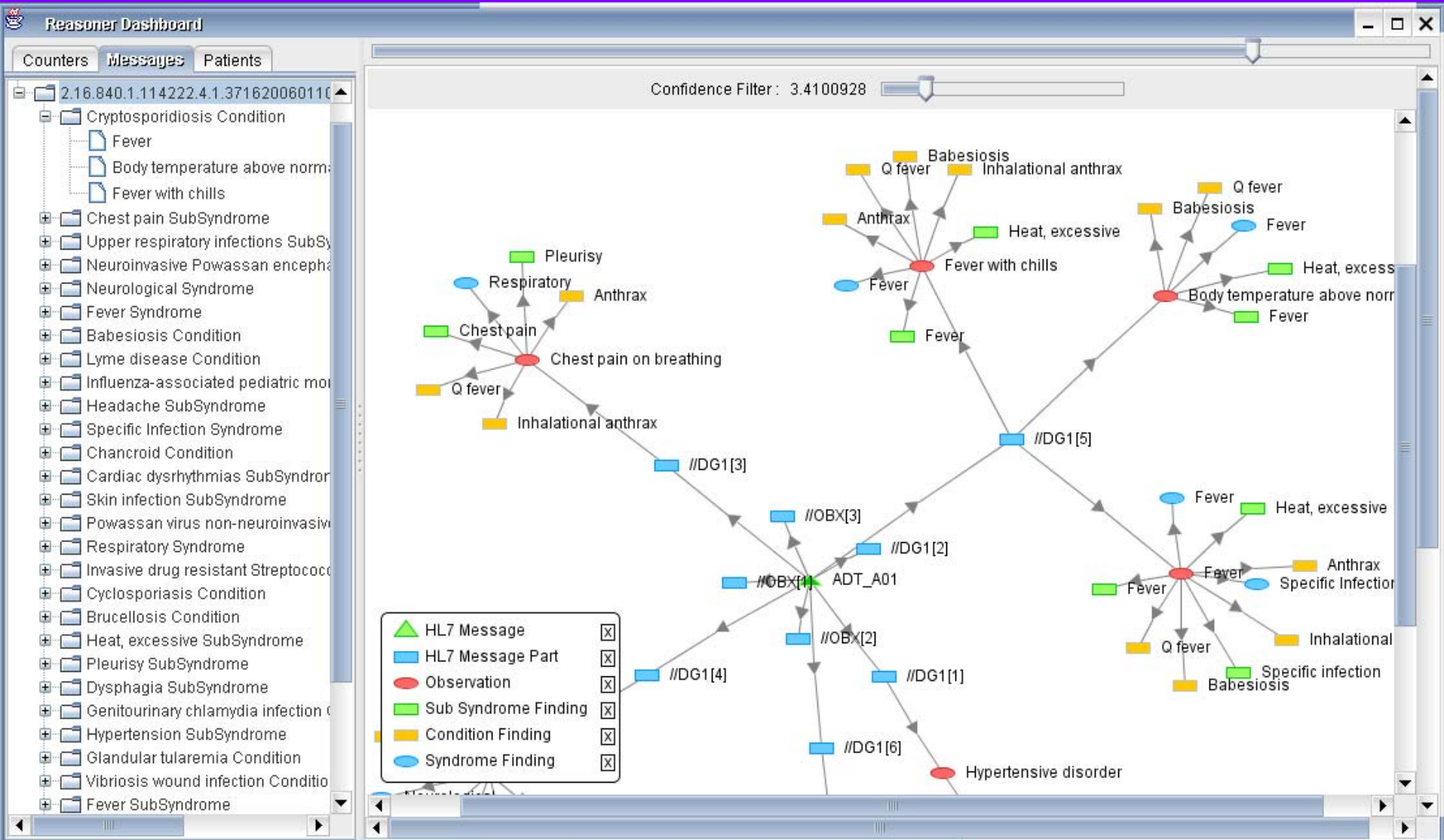
public health case [CASE]      event (occurrence) [EVN]



Close

# 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

# JESS Rule From Template

```
(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-finding (cond-code ?condCode) (finding-code ?observation)(high ?prob) (ratio
?ratio))
  (nnd-condition (cond-code ?condCode) (description ?desc))
  (not (BSFinding (msgId ?mId)(finding ?condCode)))
=>
  (assert (BSFinding (msgId ?mId)(patientId ?patientId) (findingId ?*ClassId*)
    (finding ?condCode)(findingType Condition)(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)(explanation ?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

