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

Cecil Lynch[1,2], Craig Cunningham[1], Eric Schripsema[1], Tim Morris[3], Barry Rhodes[3]

[1] OntoReason,LLC, [2] UC Davis, [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

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

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

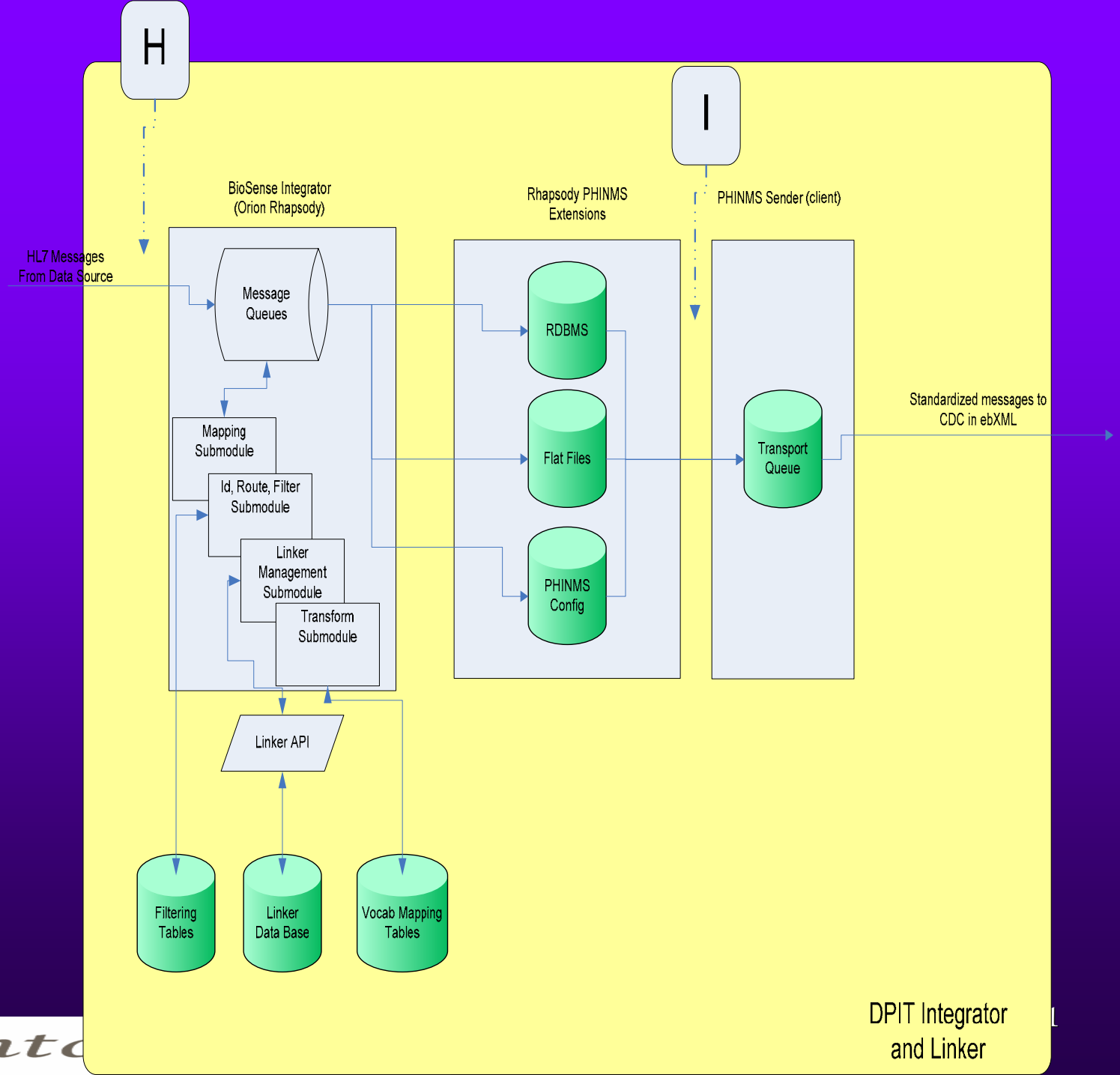
- Data mapped to 11 syndrome categories
  - Botulism-like
  - Fever
  - Gastrointestinal
  - Hemorrhagic illness
  - Localized cutaneous lesion
  - Lymphadenitis
  - Neurological
  - Rash
  - Respiratory
  - Severe illness/death
  - Specific infection
- 79 sub-syndrome categories

Watch what you ask for!

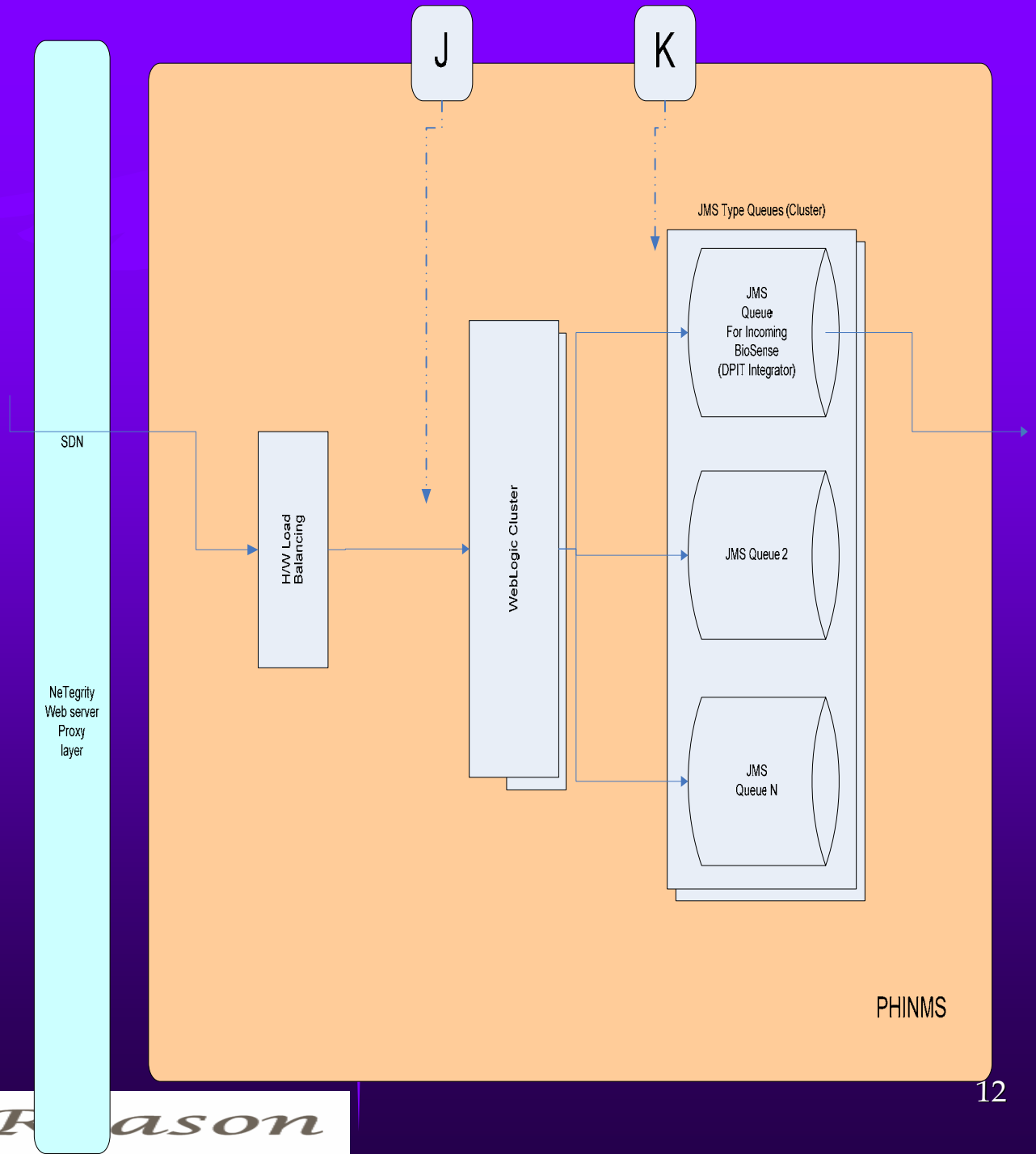
- BioSense message volume capacity today
  - 837 messages a second
  - >72 million messages a day
- How does an epidemiologist review that volume of data?
- How do you link messages to an individual over time to refine the diagnostic info?

Current BioSense Framework

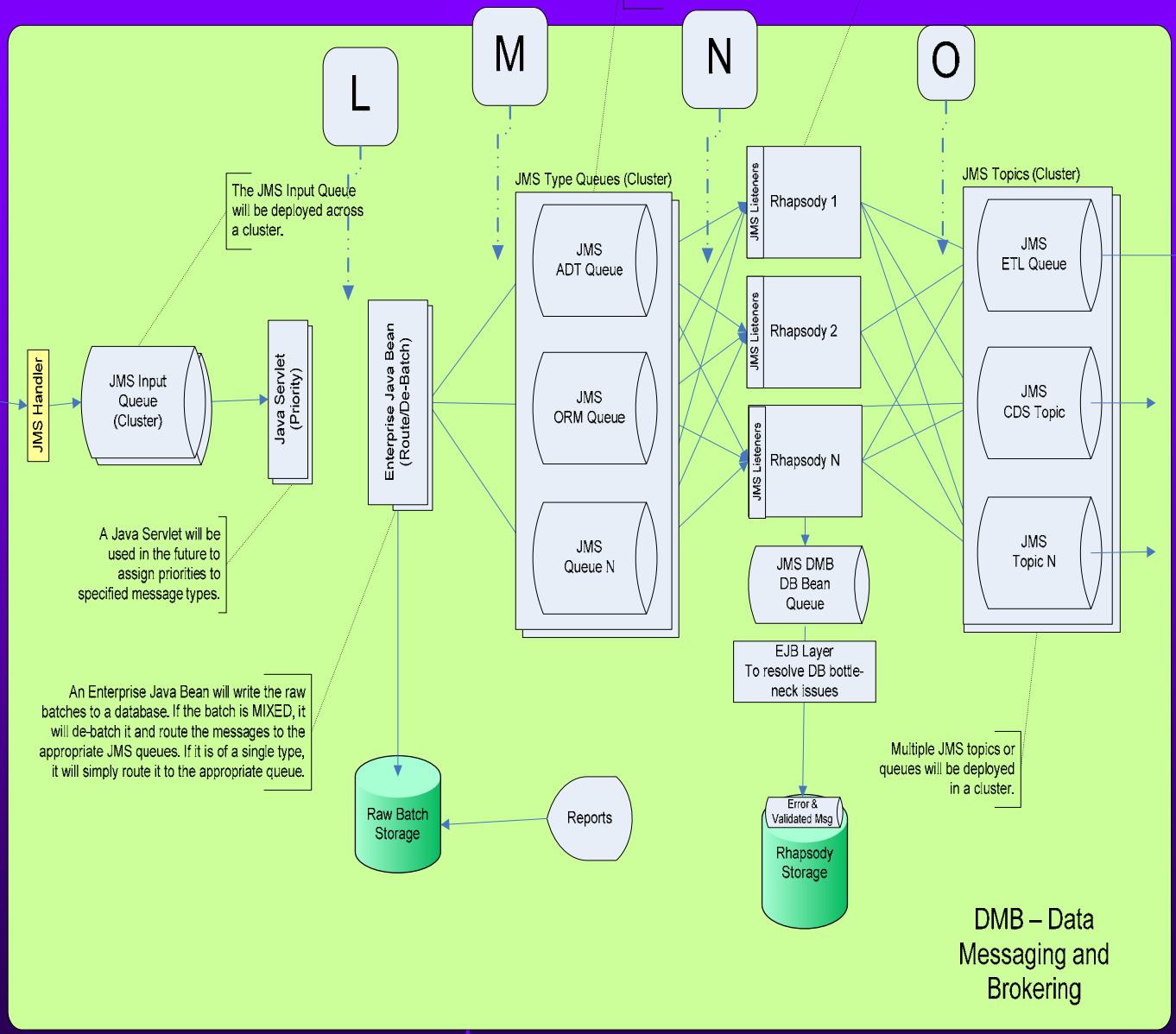
Message Processing



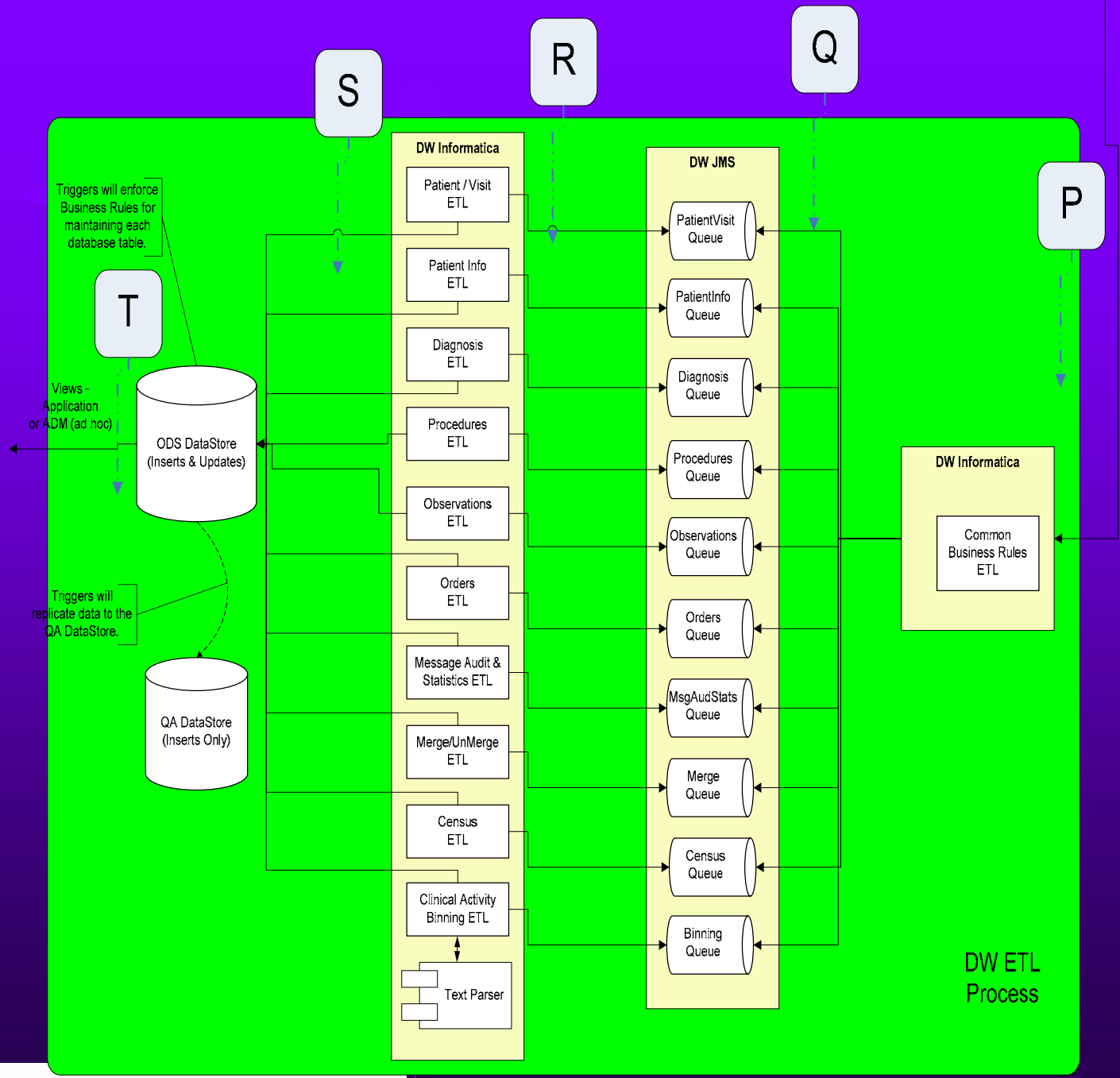
Load Balancing



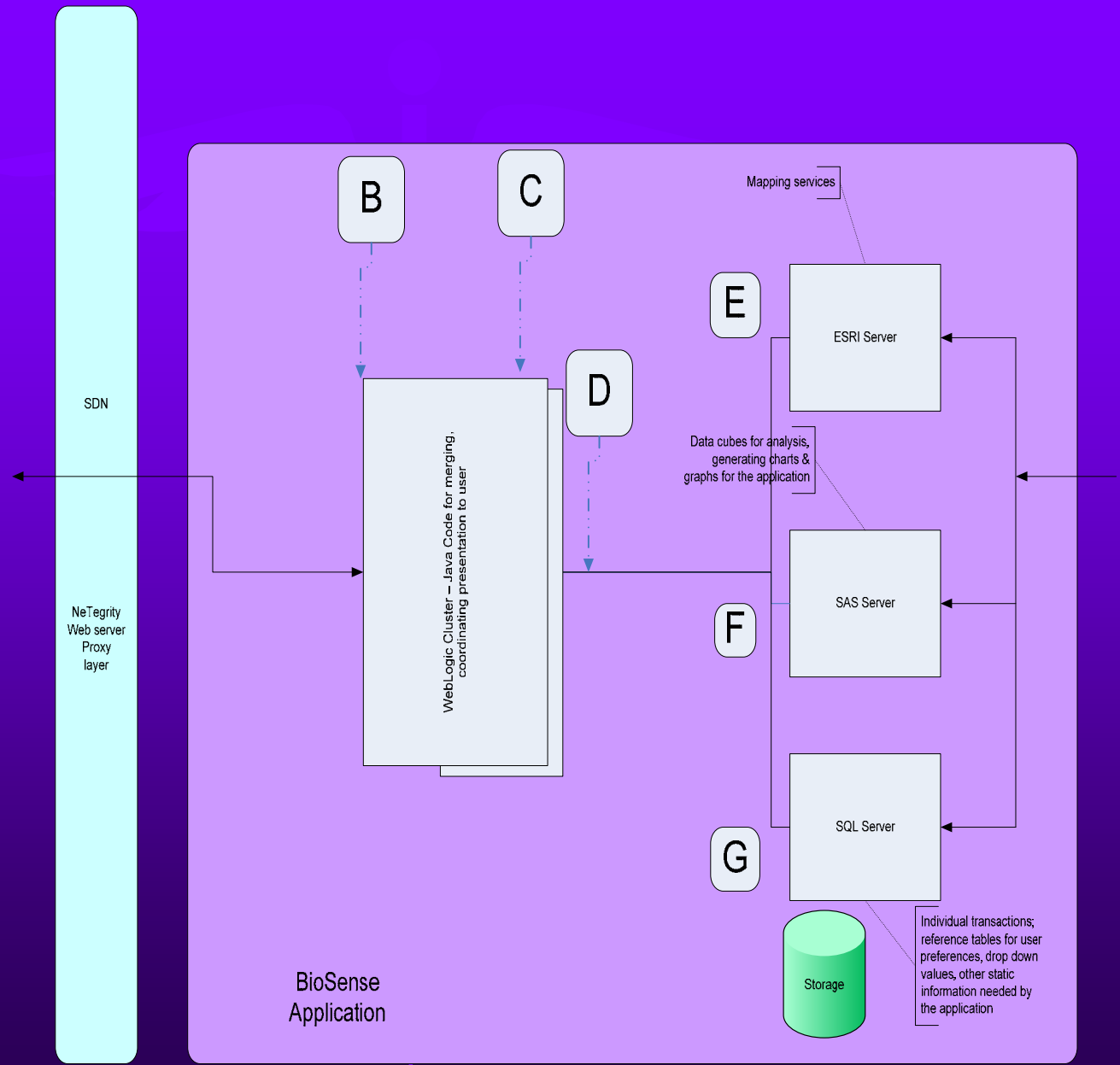
Message Type Filter



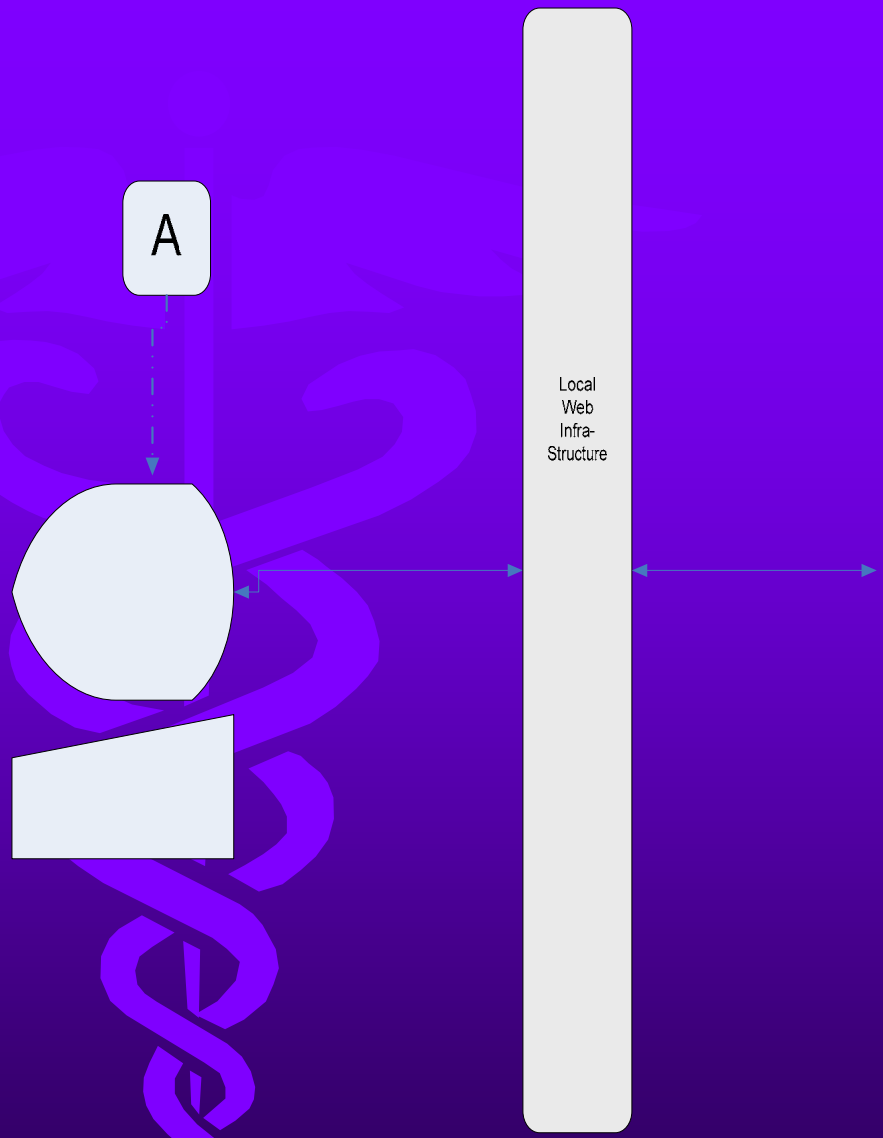
ETL Processing



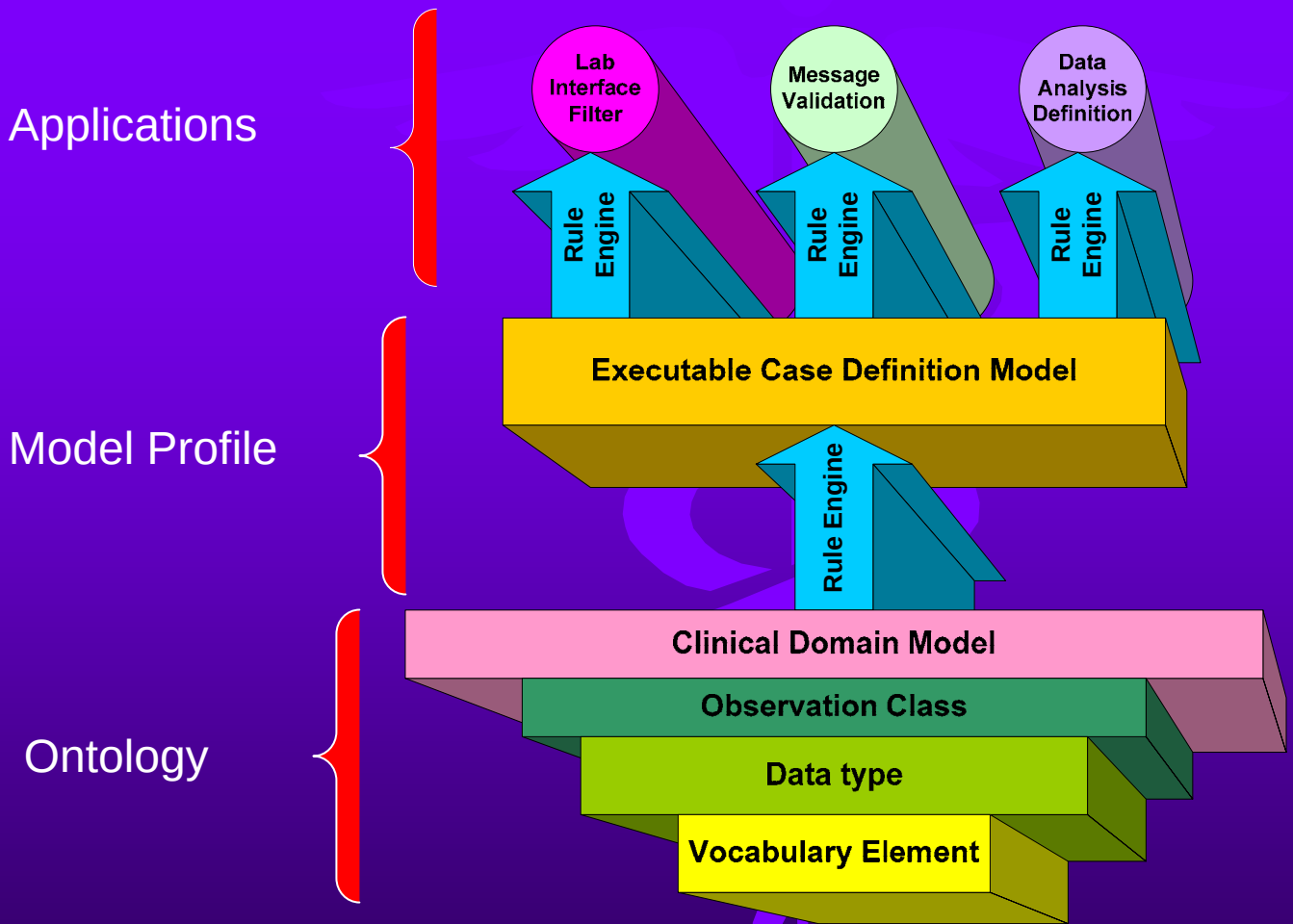
AV and OTP

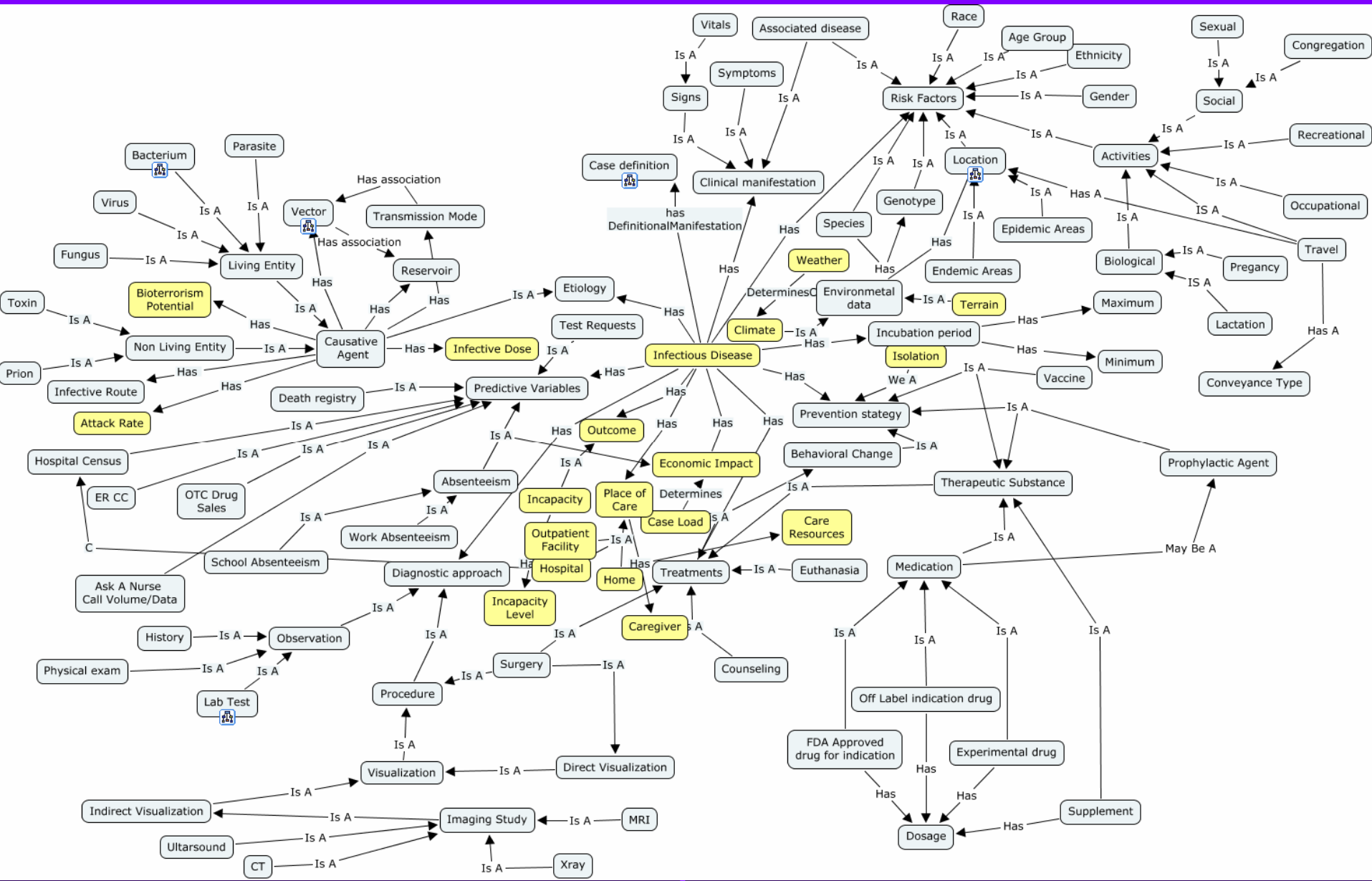

SDN
NeTegrity
Web server
Proxy
layer
B
C
D
E
F
G
Mapping services
ESRI Server
Data cubes for analysis,
generating charts &
graphs for the application
SAS Server
SQL Server
Storage
Individual transactions;
reference tables for user
preferences, drop down
values, other static
information needed by
the application
WebLogic Cluster – Java Code for merging,
coordinating presentation to user
BioSense
Application

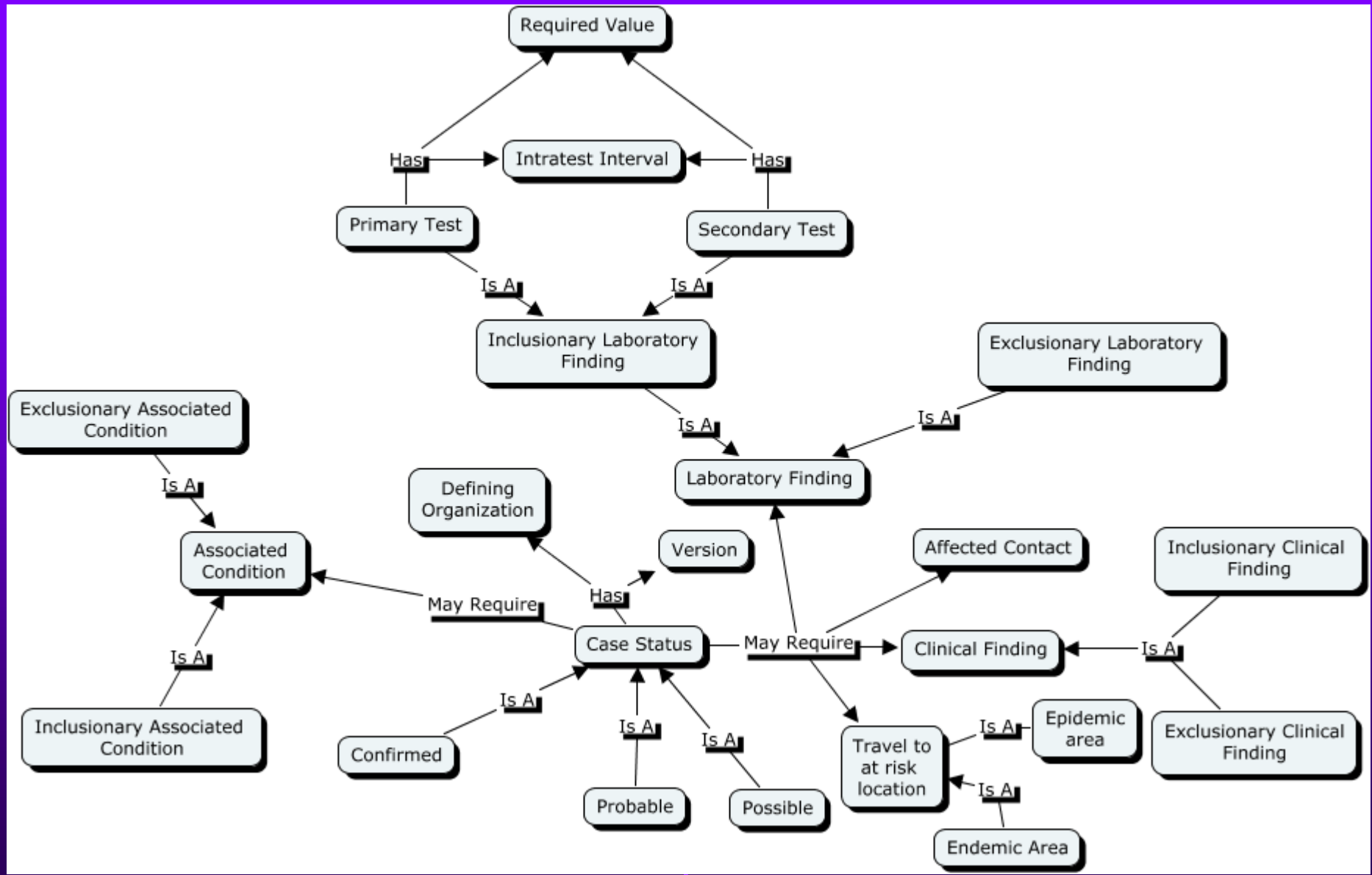
End User Views



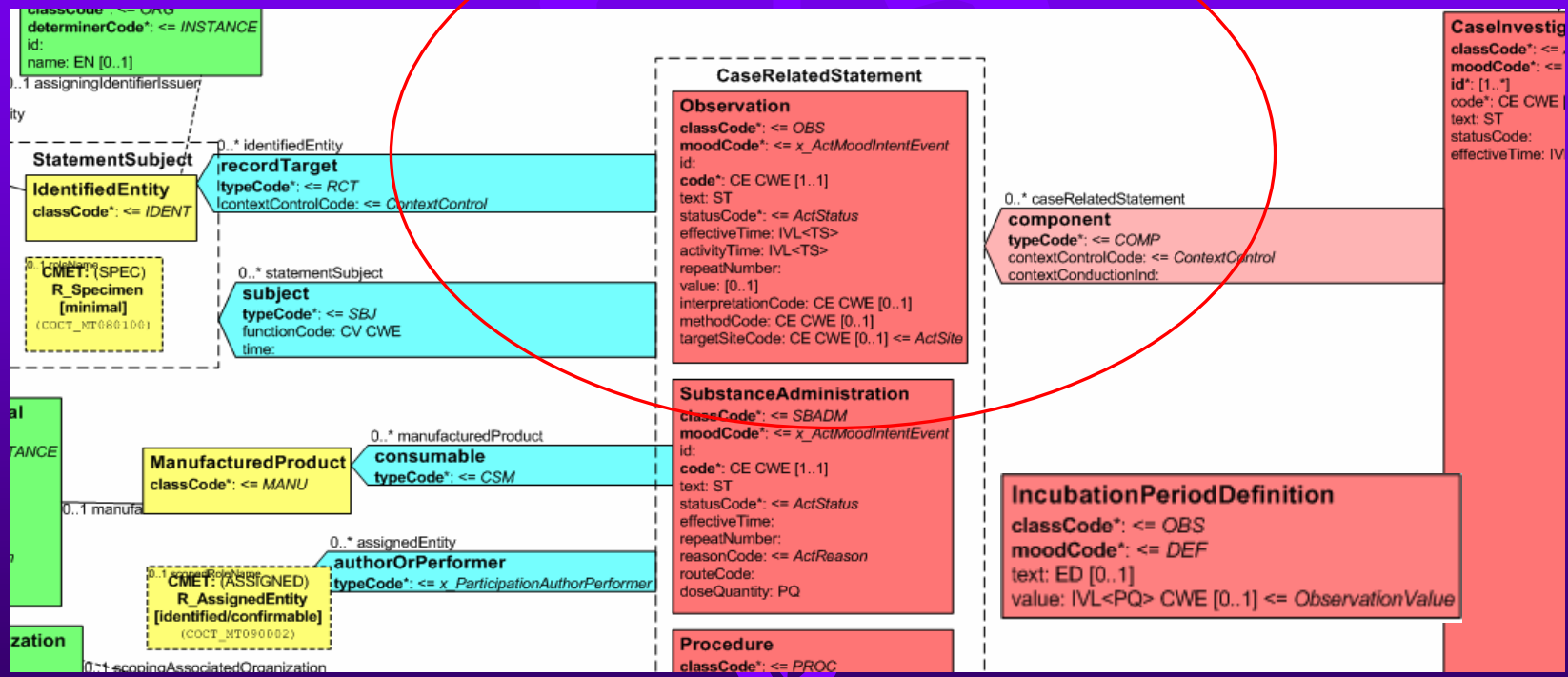
The OntoReason PH Ontology


Applications
Model Profile
Ontology
Lab Interface Filter
Message Validation
Data Analysis Definition
Rule Engine
Rule Engine
Rule Engine
Executable Case Definition Model
Rule Engine
Clinical Domain Model
Observation Class
Data type
Vocabulary Element

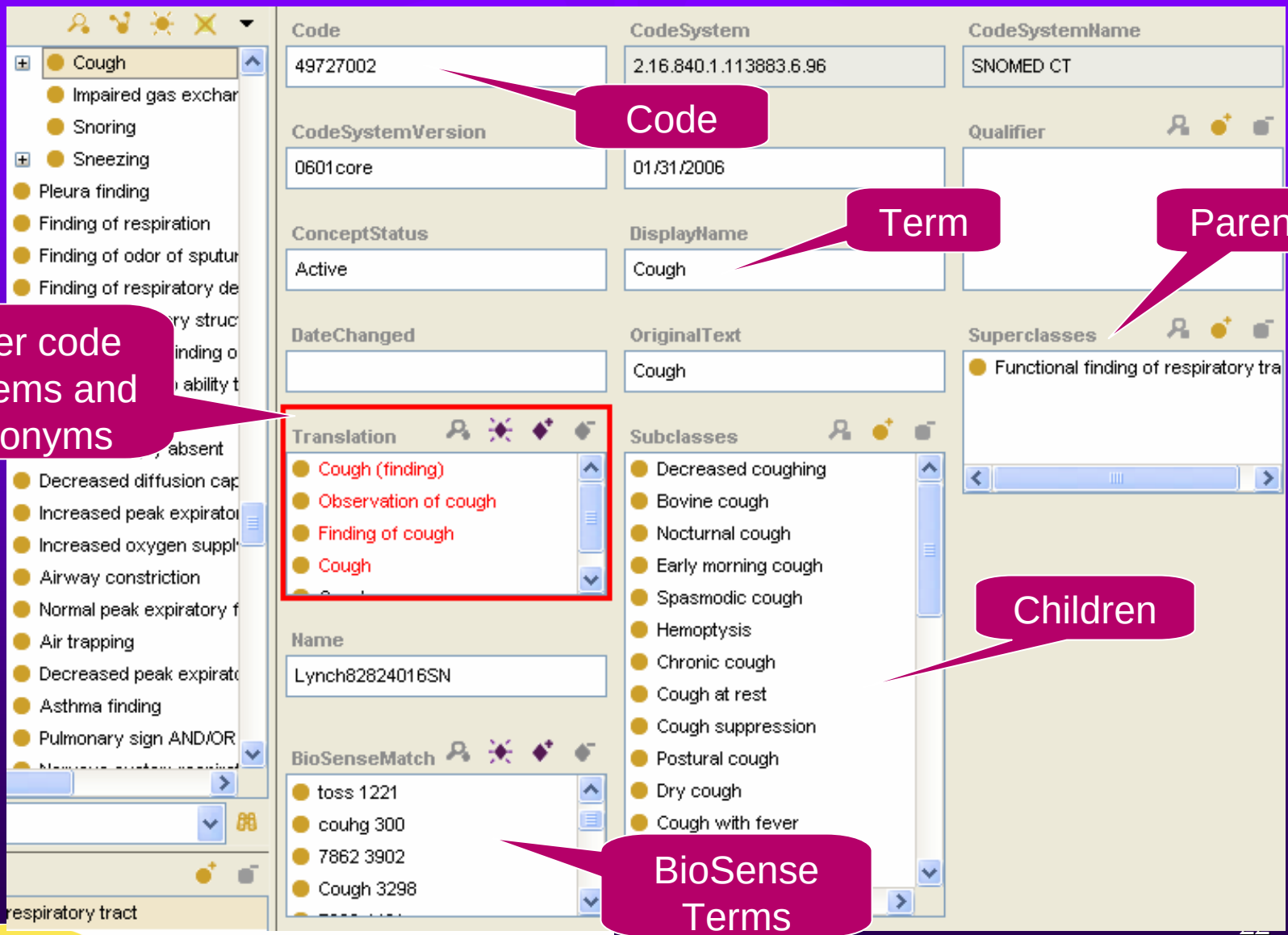

Infectious Disease
Causative Agent
Bacterium
Parasite
Virus
Fungus
Toxin
Prion
Living Entity
Non Living Entity
Vector
Reservoir
Has association
Transmission Mode
Bioterrorism Potential
Infective Route
Attack Rate
Infective Dose
Etiology
Test Requests
Predictive Variables
Death registry
Hospital Census
ER CC
OTC Drug Sales
Ask A Nurse Call Volume/Data
School Absenteeism
Work Absenteeism
Absenteeism
Incapacity
Incapacity Level
Outpatient Facility
Hospital
Home
Place of Care
Case Load
Outcome
Economic Impact
Diagnostic approach
Observation
History
Physical exam
Lab Test
Procedure
Surgery
Caregiver
Counseling
Treatments
Euthanasia
Care Resources
Visualization
Direct Visualization
Indirect Visualization
Imaging Study
MRI
Ultarsound
CT
Xray
Vitals
Signs
Symptoms
Associated disease
Clinical manifestation
Case definition
has DefinitionalManifestation
Risk Factors
Race
Age Group
Ethnicity
Gender
Genotype
Species
Weather
Climate
Determines
Environmetal data
Terrain
Location
Epidemic Areas
Endemic Areas
Activities
Sexual
Social
Congregation
Recreational
Occupational
Travel
Biological
Pregancy
Lactation
Conveyance Type
Has A
Incubation period
Maximum
Minimum
Isolation
Vaccine
We A
Prevention stategy
Behavioral Change
Therapeutic Substance
Prophylactic Agent
May Be A
Medication
Off Label indication drug
FDA Approved drug for indication
Experimental drug
Dosage
Supplement
Has
Is A


Required Value
Has
Intratest Interval
Has
Primary Test
Secondary Test
Is A
Is A
Inclusionary Laboratory Finding
Exclusionary Laboratory Finding
Is A
Exclusionary Associated Condition
Is A
Laboratory Finding
Defining Organization
Version
Affected Contact
Inclusionary Clinical Finding
Associated Condition
May Require
Has
Case Status
May Require
Clinical Finding
Is A
Is A
Is A
Inclusionary Associated Condition
Confirmed
Is A
Is A
Is A
Probable
Possible
Travel to at risk location
Is A
Epidemic area
Exclusionary Clinical Finding
Is A
Endemic Area

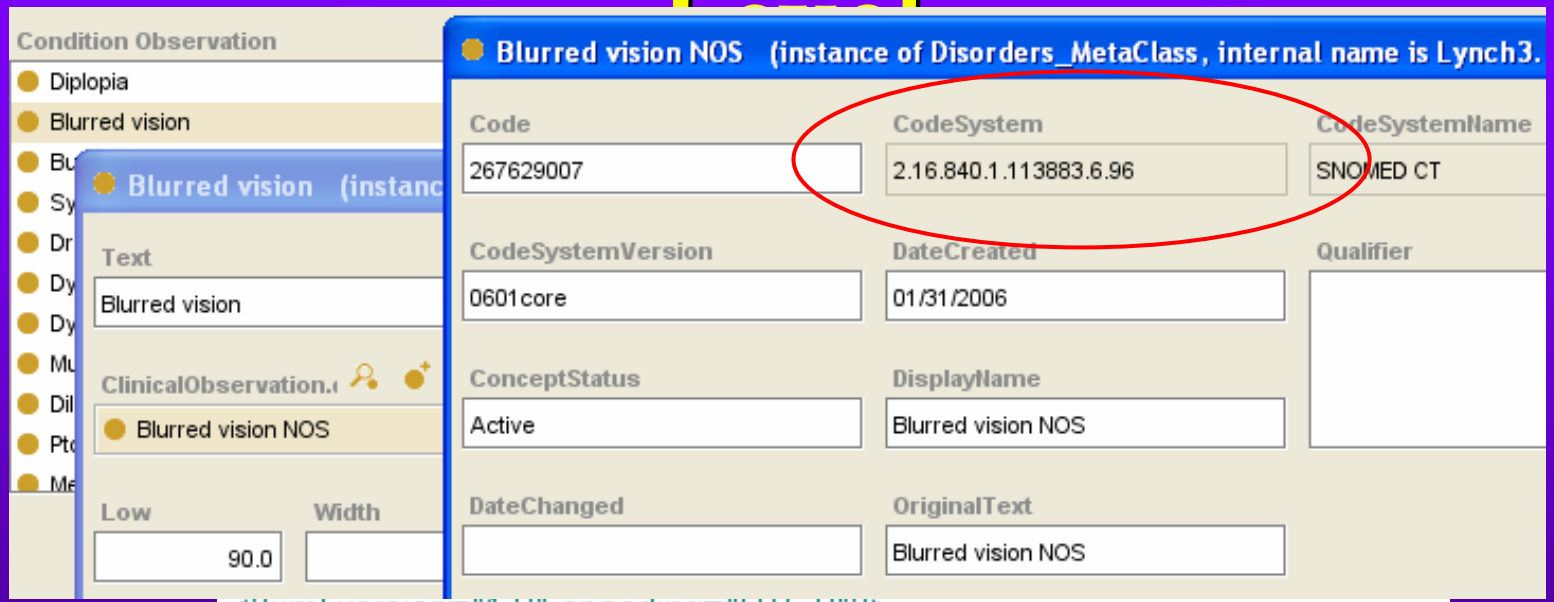
Information Model



Concept In HL7 V3 DataType

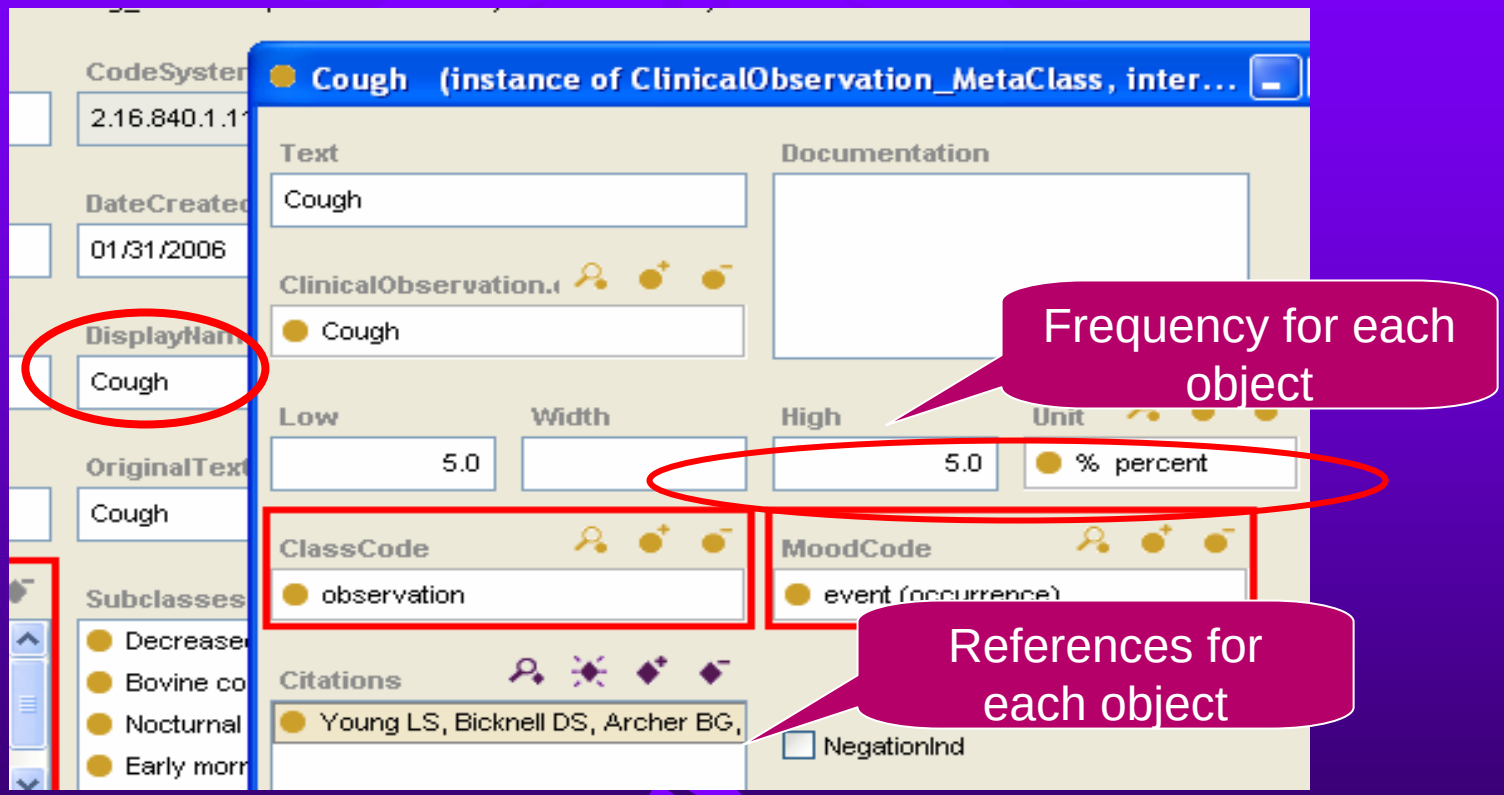


Conceptual and Syntactical Level

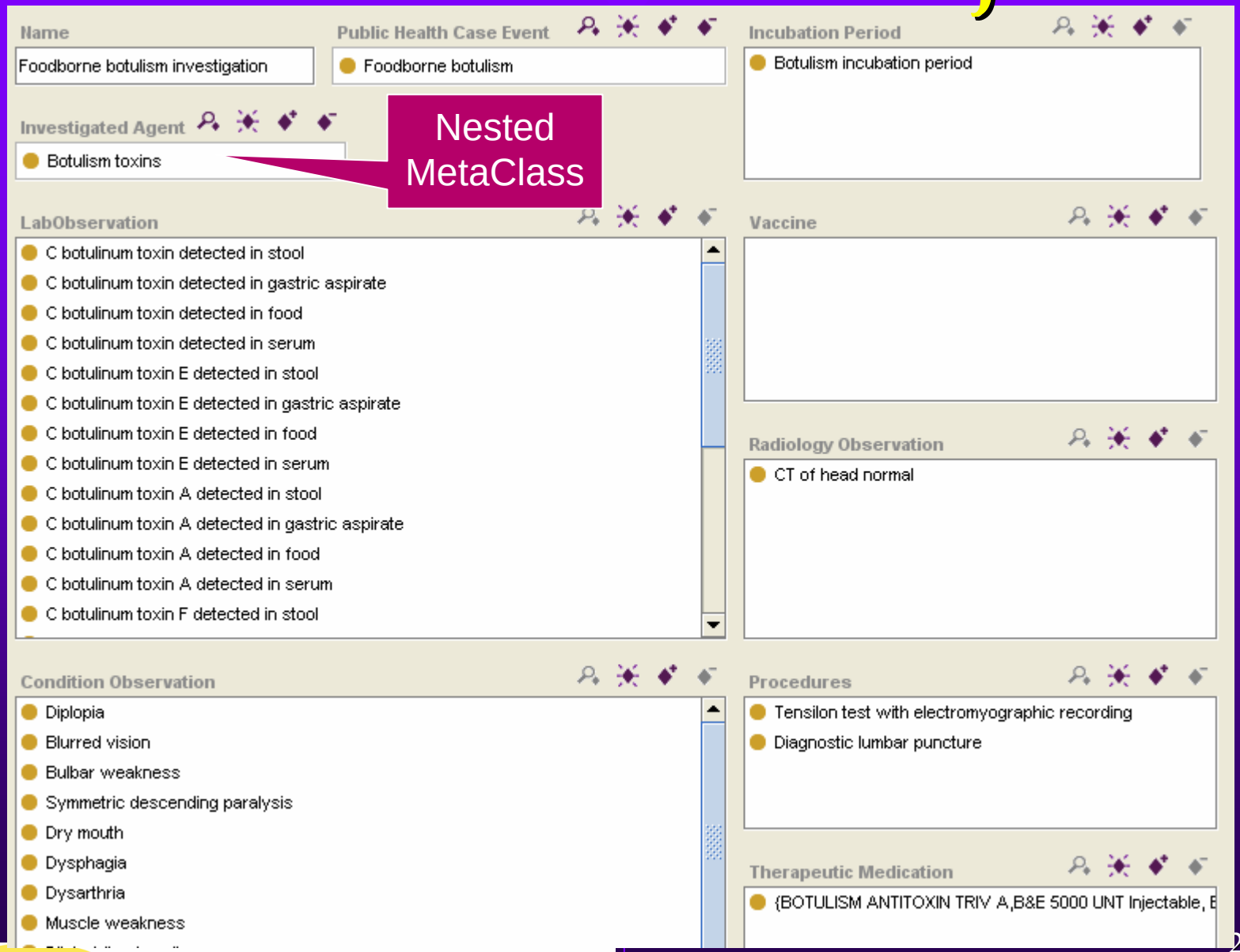


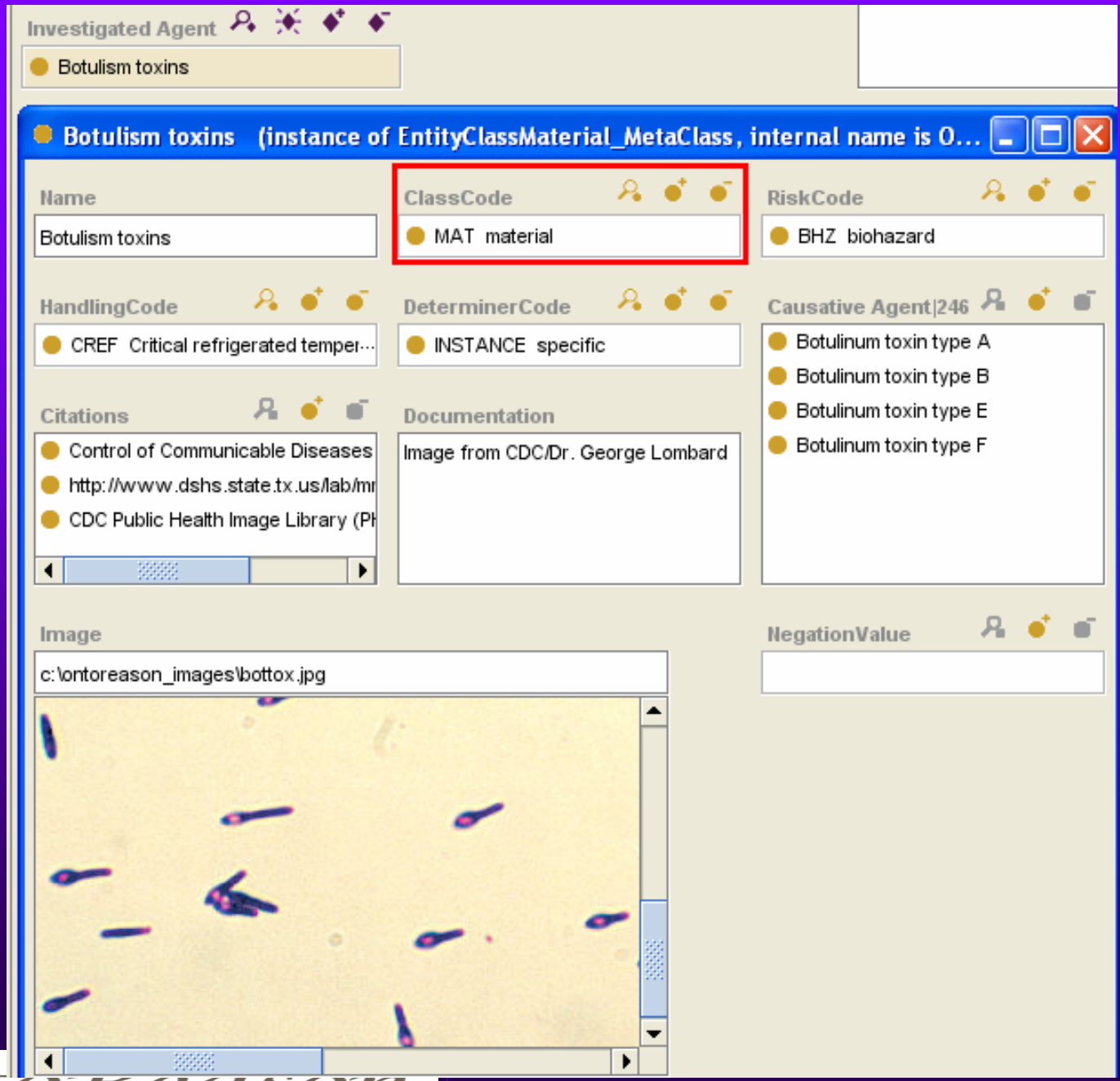

```
<?xml version="1.0" encoding="UTF-8"?>
< xmlns:hl7v3=" http://www.w3.org/TR/WD-xlink" xmlns:hl7="urn:hl7-org:v3"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"/>

<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

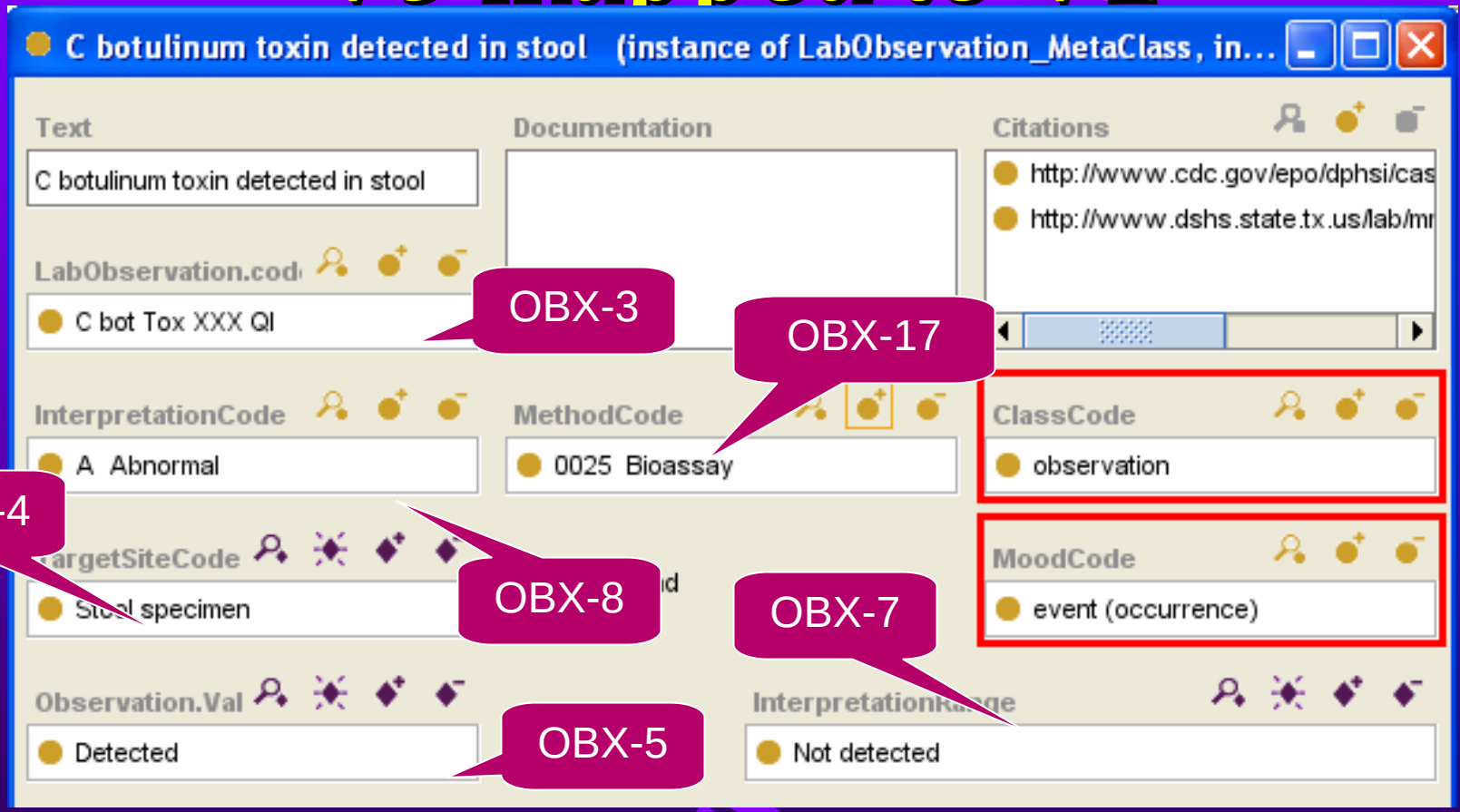


Clinical Domain Object

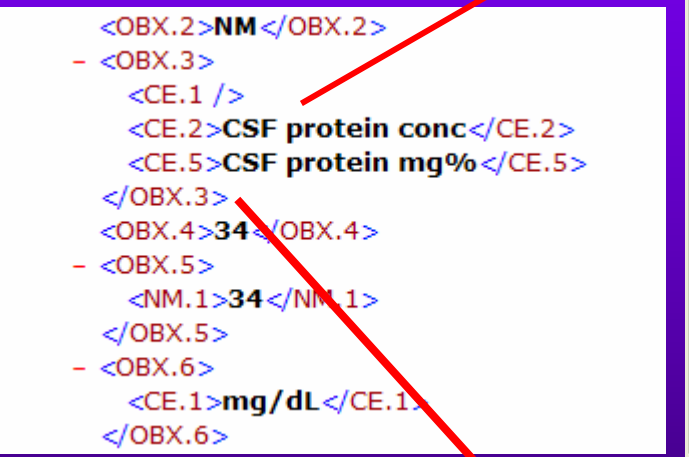


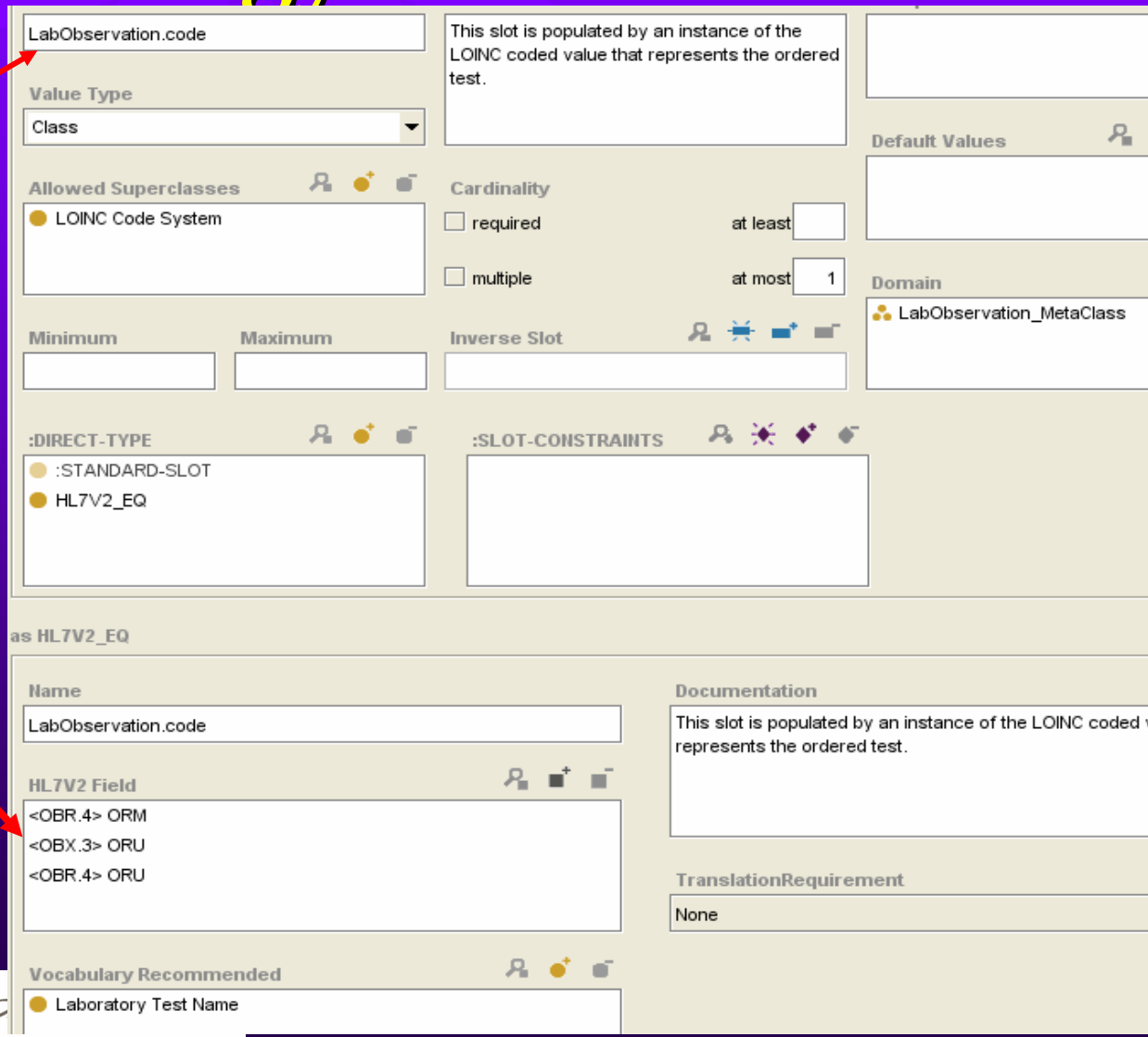

Investigated Agent
Botulism toxins
Botulism toxins (instance of EntityClassMaterial_MetaClass, internal name is O...
Name
Botulism toxins
ClassCode
MAT material
RiskCode
BHZ biohazard
HandlingCode
CREF Critical refrigerated temper...
DeterminerCode
INSTANCE specific
Causative Agent|246
Botulinum toxin 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/mr
CDC Public Health Image Library (PH
Documentation
Image from CDC/Dr. George Lombard
Image
c:\ontoreason_images\bottox.jpg
NegationValue

Laboratory Observation HL7 V3 mapped to V2



Map HL7 Message segments to Ontology Slots





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

OntoReason

Ontology Extraction

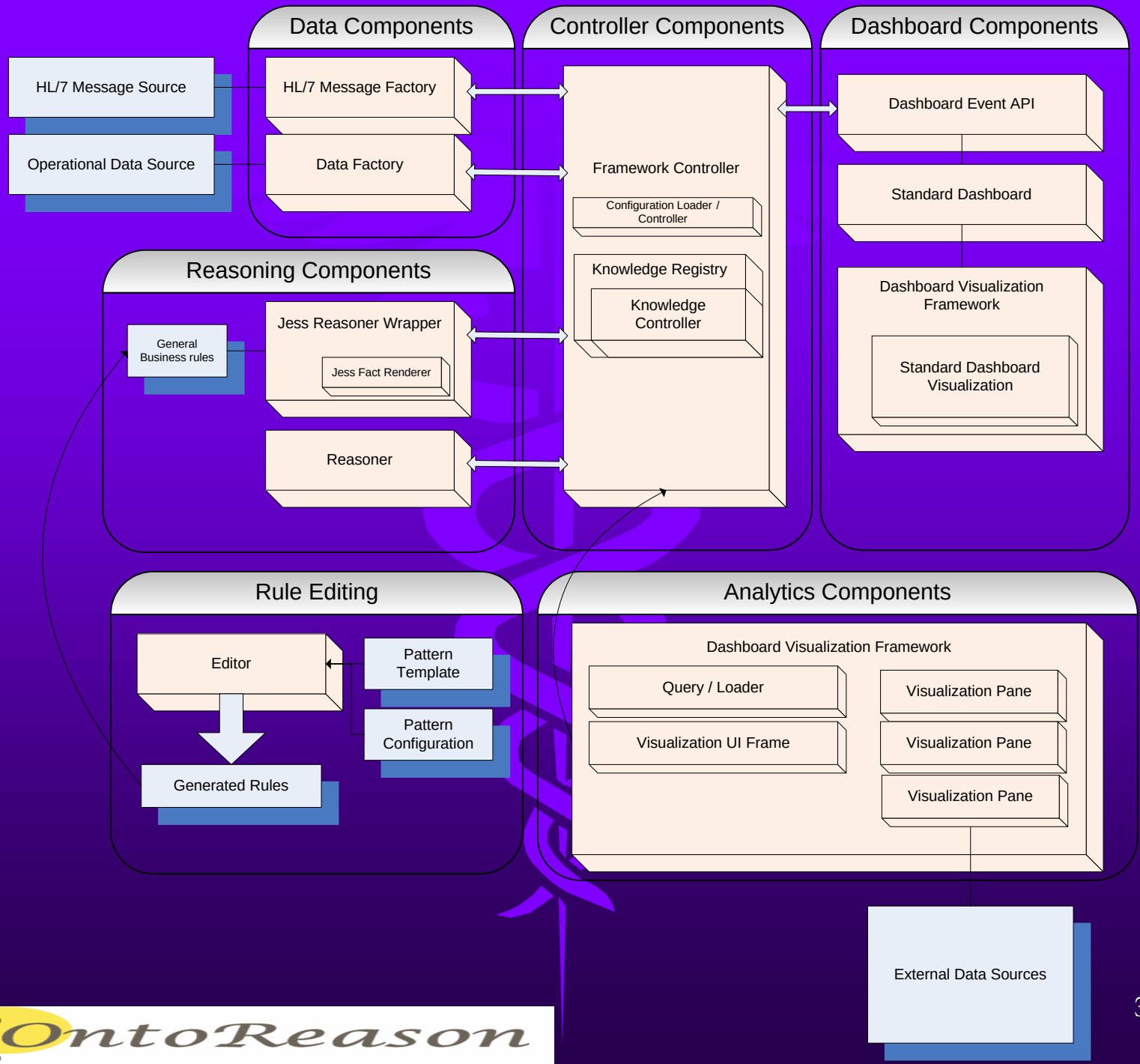
- **Creating an application ontology from the reference ontology**
  - Identify the core ontology classes
  - Create an object representation that maintains the ontology data
  - Generate cross reference indexes for core relationships
    - Lab tests to case investigations
    - Organism/Agent to case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

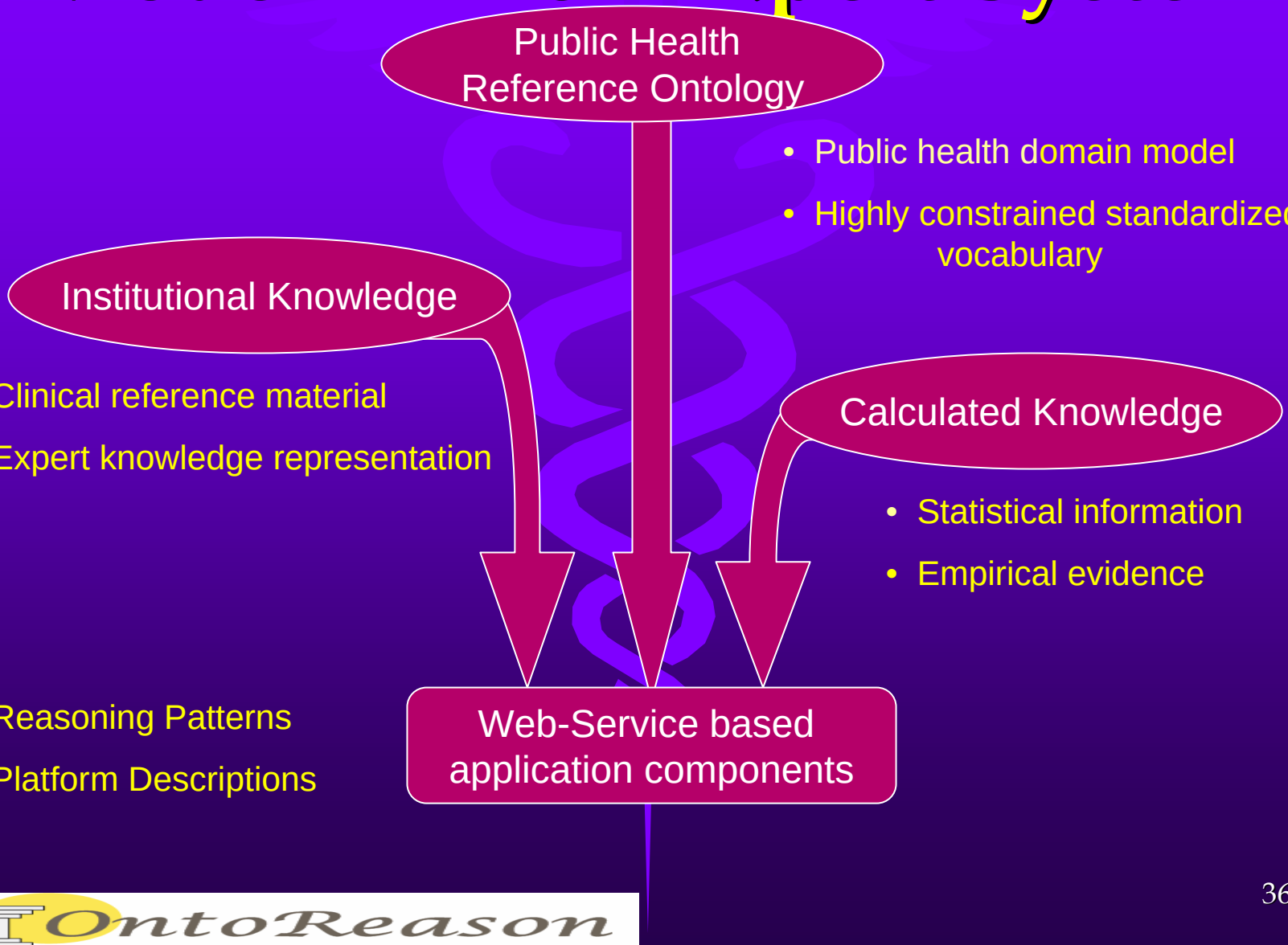
Ontology Representation

- **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 on 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


Data Components
Controller Components
Dashboard Components
HL/7 Message Source
Operational Data Source
HL/7 Message Factory
Data Factory
Framework Controller
Configuration Loader / Controller
Knowledge Registry
Knowledge Controller
Dashboard Event API
Standard Dashboard
Dashboard Visualization Framework
Standard Dashboard Visualization
Reasoning Components
General Business rules
Jess Reasoner Wrapper
Jess Fact Renderer
Reasoner
Rule Editing
Editor
Pattern Template
Pattern Configuration
Generated Rules
Analytics Components
Dashboard Visualization Framework
Query / Loader
Visualization UI Frame
Visualization Pane
Visualization Pane
Visualization Pane
External Data Sources

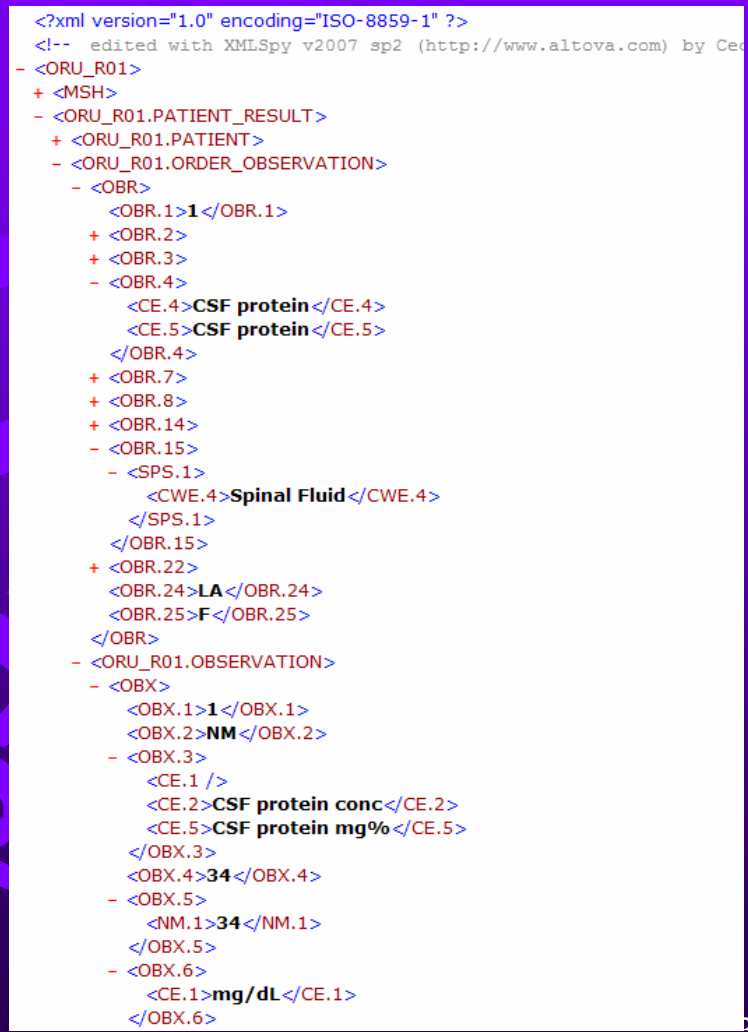
Message Analytics Workbench

Model Driven Expert System



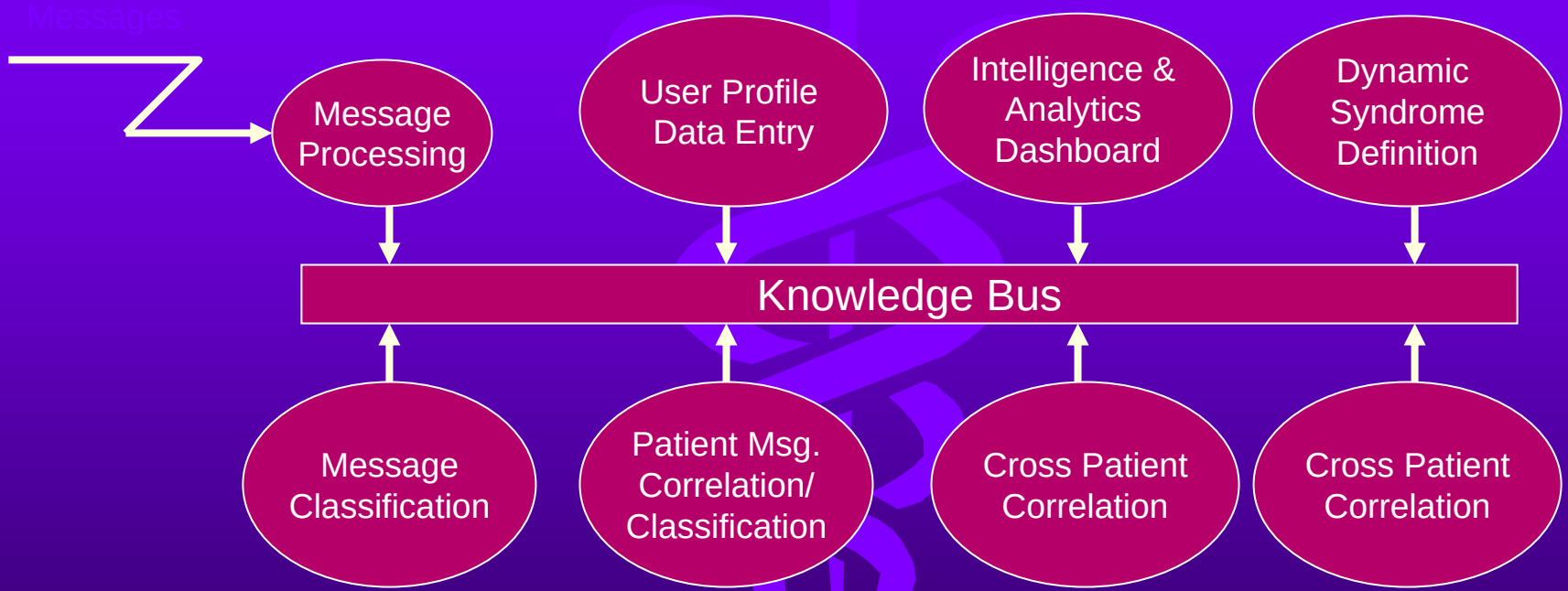
BioSense Message Data Source

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

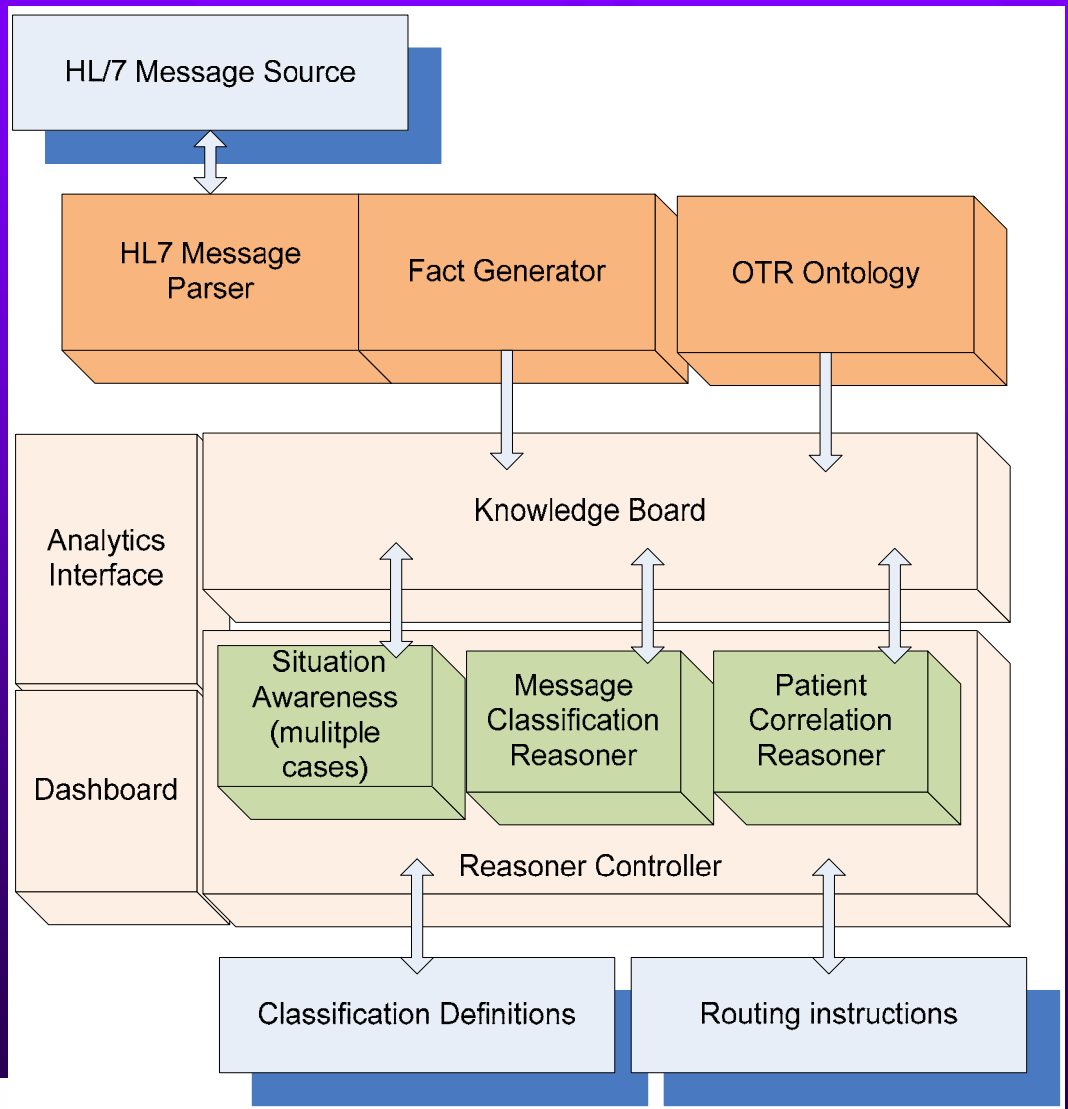


```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- edited with XMLSpy v2007 sp2 (http://www.altova.com) by Ce
- <ORU_R01>
  + <MSH>
  - <ORU_R01.PATIENT_RESULT>
    + <ORU_R01.PATIENT>
    - <ORU_R01.ORDER_OBSERVATION>
      - <OBR>
          <OBR.1>1</OBR.1>
        + <OBR.2>
        + <OBR.3>
        - <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?

- Anthrax investigation (inhalation)
  - 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
  - Organisims/Agents
    - **Bacillus anthracis** (21927003)
  - Radiology Observations

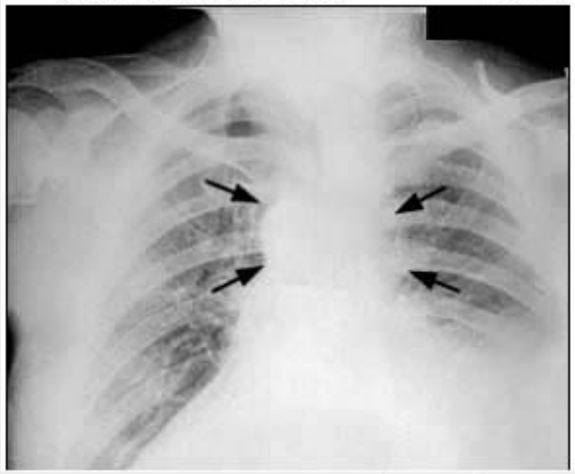
Anthrax investigation (inhalation)

Citations

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

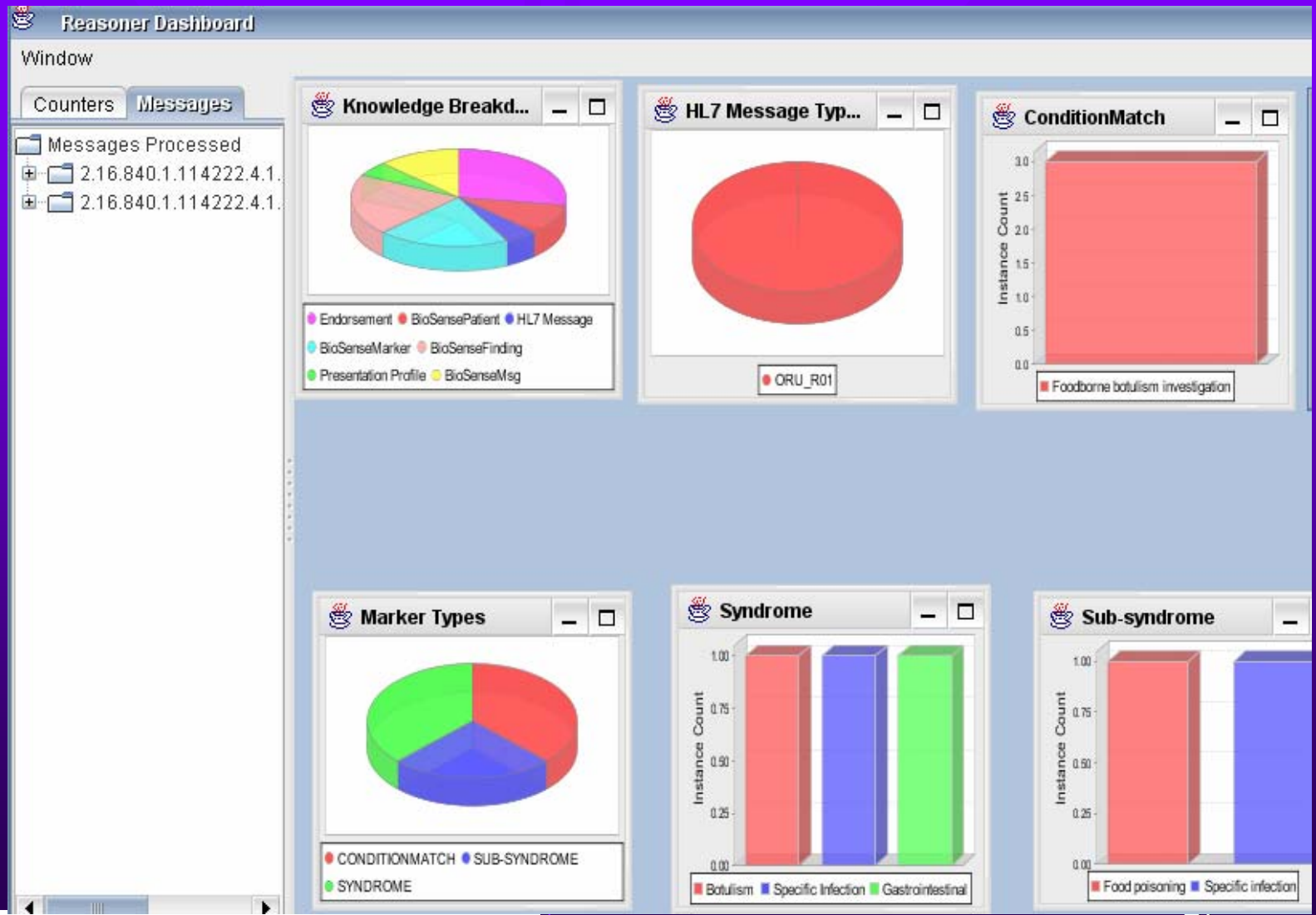
HL7Representation

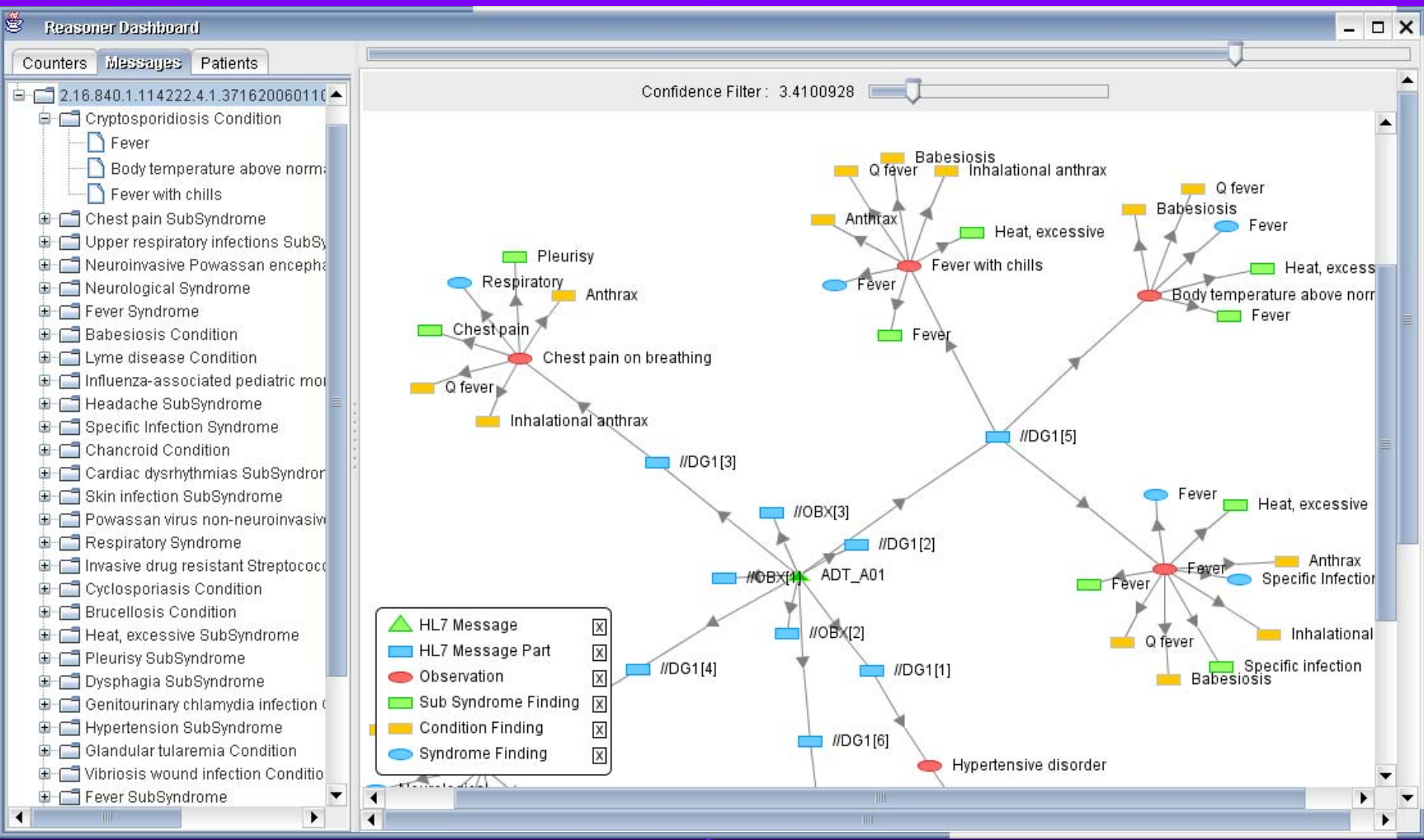
public health case [CASE] event (occurrence) [EVN]



Close

Reasoner results




Reasoner Dashboard
Counters
Messages
Patients
2.16.840.1.114222.4.1.371620060110
Cryptosporidiosis Condition
Fever
Body temperature above norm
Fever with chills
Chest pain SubSyndrome
Upper respiratory infections SubSy
Neuroinvasive Powassan encepha
Neurological Syndrome
Fever Syndrome
Babesiosis Condition
Lyme disease Condition
Influenza-associated pediatric mor
Headache SubSyndrome
Specific Infection Syndrome
Chancroid Condition
Cardiac dysrhythmias SubSyndror
Skin infection SubSyndrome
Powassan virus non-neuroinvasiv
Respiratory Syndrome
Invasive drug resistant Streptococ
Cyclosporiasis Condition
Brucellosis Condition
Heat, excessive SubSyndrome
Pleurisy SubSyndrome
Dysphagia SubSyndrome
Genitourinary chlamydia infection
Hypertension SubSyndrome
Glandular tularemia Condition
Vibriosis wound infection Conditio
Fever SubSyndrome
Confidence Filter : 3.4100928
Babesiosis
Q fever
Inhalational anthrax
Anthrax
Heat, excessive
Fever with chills
Fever
Fever
Pleurisy
Respiratory
Anthrax
Chest pain
Chest pain on breathing
Q fever
Inhalational anthrax
Q fever
Babesiosis
Fever
Heat, excess
Body temperature above norr
Fever
//DG1[5]
//DG1[3]
//OBX[3]
//DG1[2]
//OBX[4]
ADT_A01
//OBX[2]
//DG1[4]
//DG1[1]
//DG1[6]
Hypertensive disorder
Fever
Heat, excessive
Anthrax
Fever
Fever
Specific Infectior
Inhalational
Q fever
Specific infection
Babesiosis
HL7 Message
HL7 Message Part
Observation
Sub Syndrome Finding
Condition Finding
Syndrome Finding

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)(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