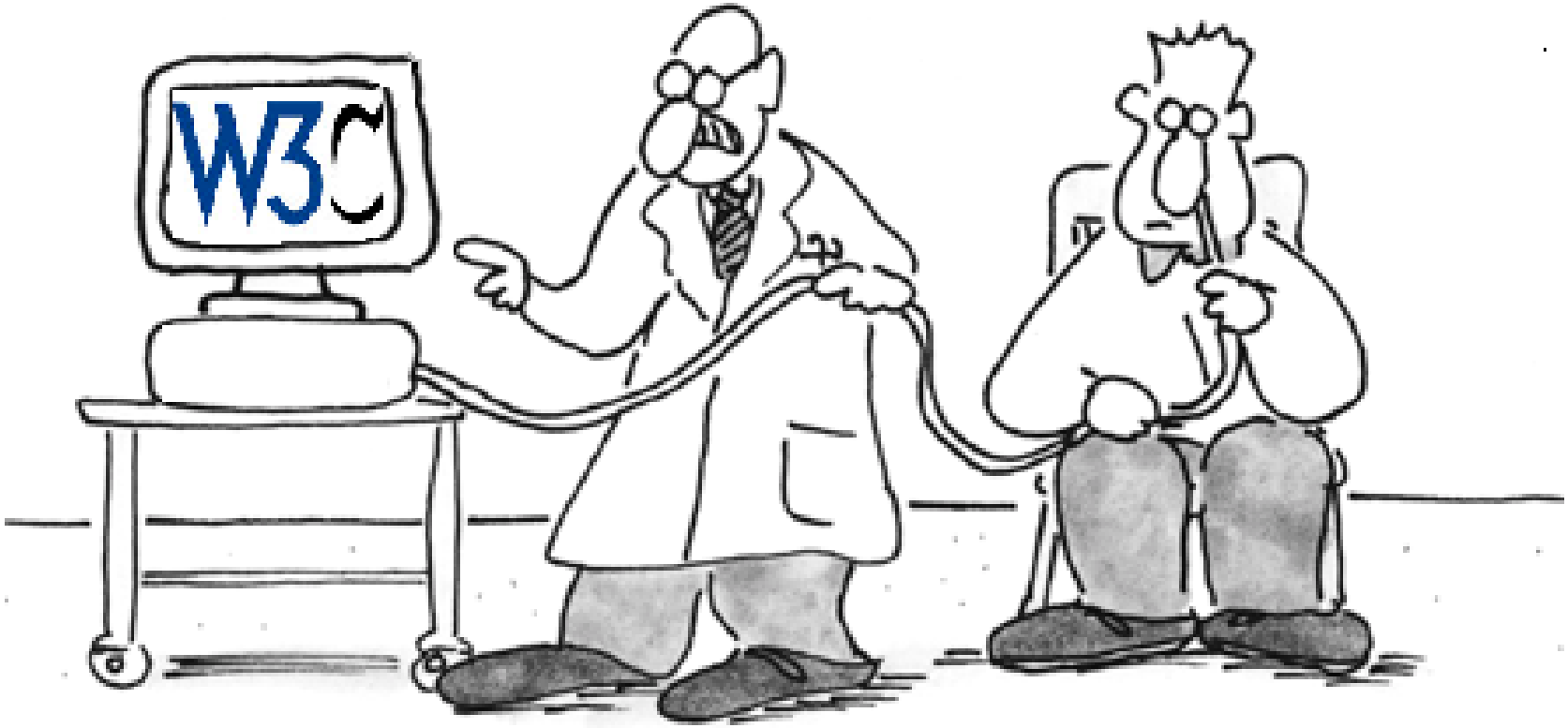
A Semantic Web Approach to Integrative Biosurveillance



Narendra Kunapareddy, UTHSC
Zhe Wu, Ph.D., Oracle

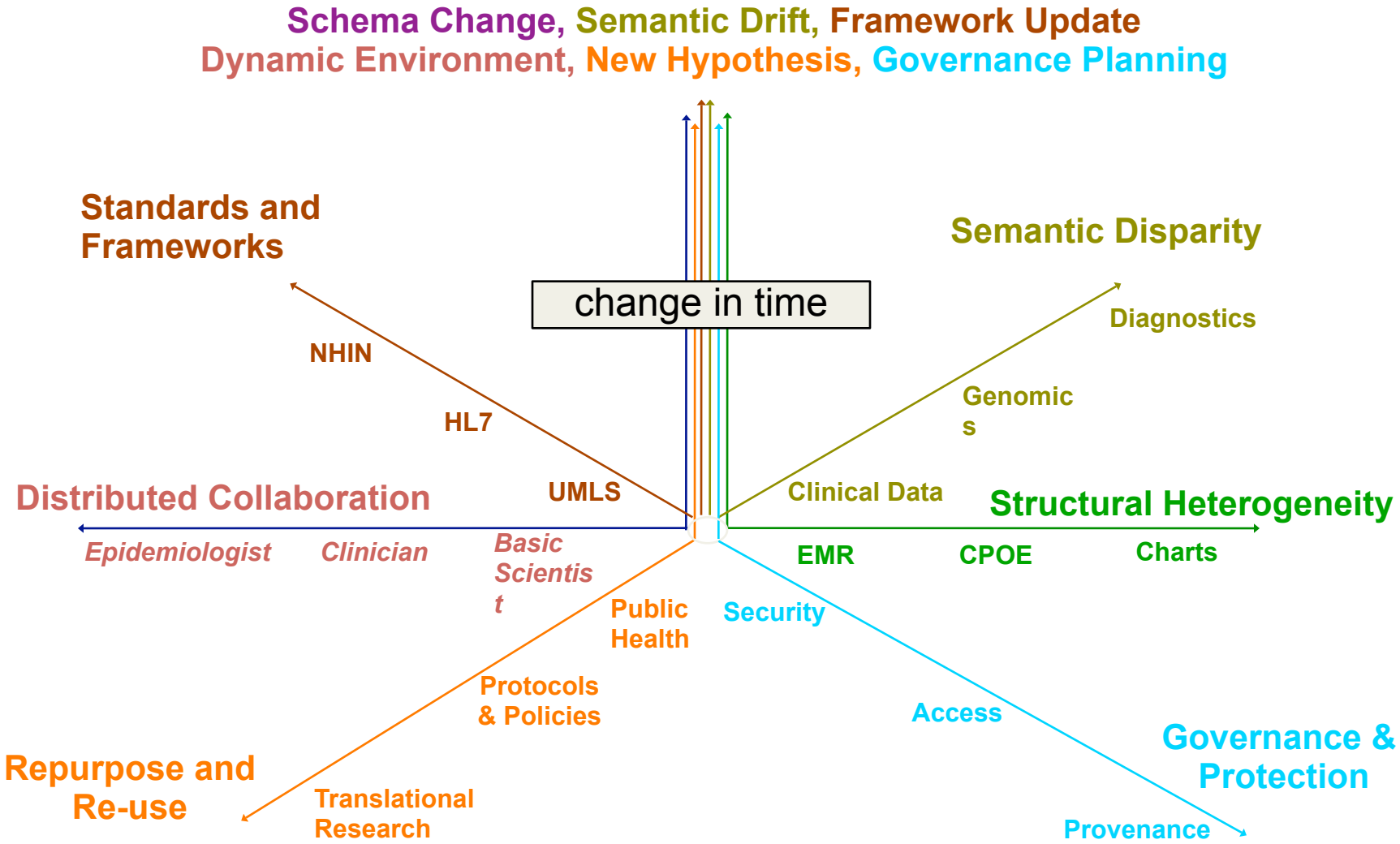
This talk:

- Translational BioInformatics and Information Integration Dilemma
- Case Study: Public Health Preparedness
- Our Vision
- Our Implementation
- Challenges

This talk:

- Translational BioInformatics and Information Integration Dilemma
- Case Study: Public Health Preparedness
- Our Vision
- Our Implementation
- Challenges

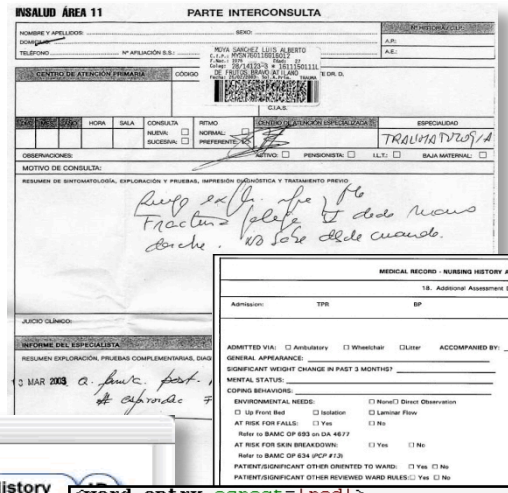
Information Integration Dilemma

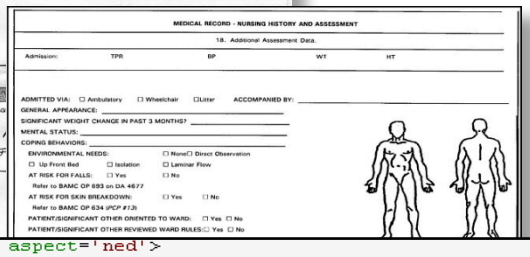


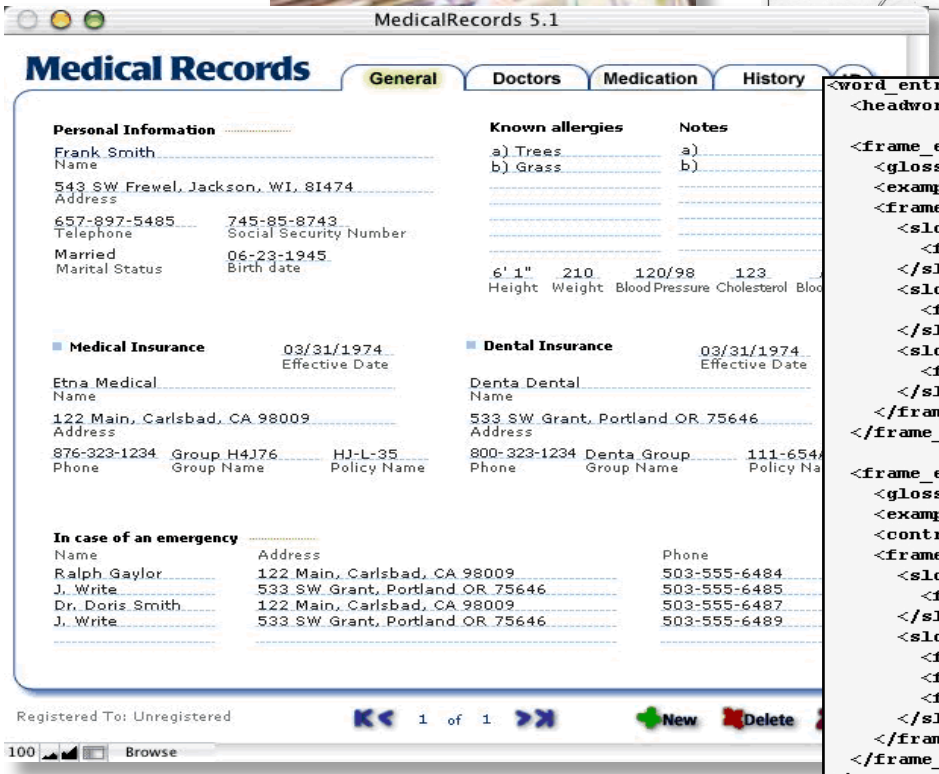
This talk:

- Translational BioInformatics and Information Integration Dilemma
- Case Study: Public Health Preparedness
- Our Vision
- Our Implementation
- Challenges

Public Health Preparedness





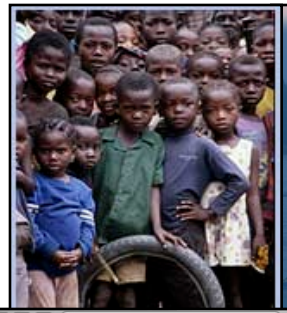


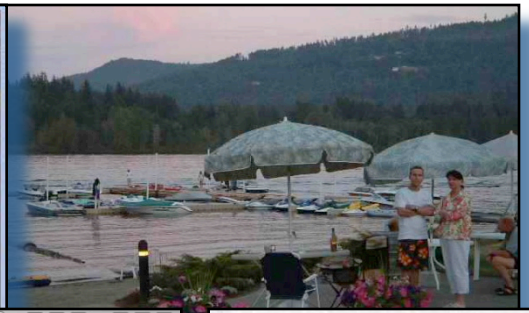

```
<word_entry aspect='ned'>
  <headword_lemma>bavit</headword_lemma>

  <frame_entry frame_index='1'>
    <gloss>poskytovat zábavu</gloss>
    <example>bavit přátele svým vyprávěním</example>
    <frame_slots>
      <slot functor='ACT' type='obl'>
        <form type='direct_case' case='1'/>
      </slot>
      <slot functor='PAT' type='obl'>
        <form type='direct_case' case='4'/>
      </slot>
      <slot functor='MEANS' type='typ'>
        <form type='direct_case' case='7'/>
      </slot>
    </frame_slots>
  </frame_entry>

  <frame_entry frame_index='2'>
    <gloss>zajímat</gloss>
    <example>to mě baví; ty mě nebavíš</example>
    <control>ACT</control>
    <frame_slots>
      <slot functor='ACT' type='obl'>
        <form type='direct_case' case='4'/>
      </slot>
      <slot functor='PAT' type='obl'>
        <form type='direct_case' case='1'/>
        <form type='infinitive'/>
        <form type='subord_conj' subord_conj_lemma='že'/>
      </slot>
    </frame_slots>
  </frame_entry>
</word_entry>
```

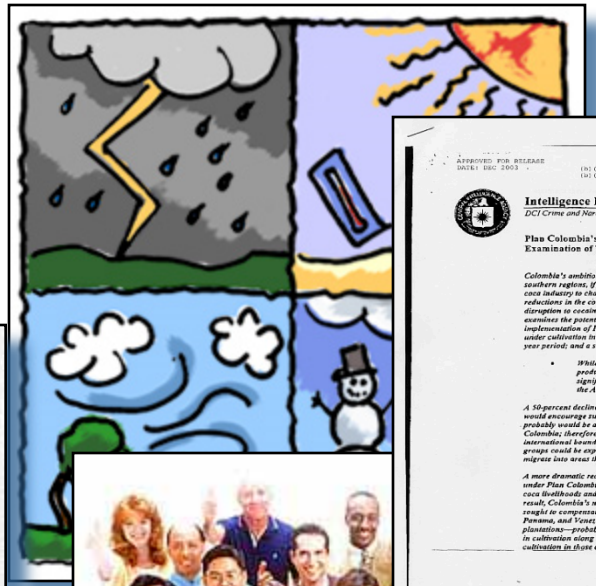
Public Health Preparedness

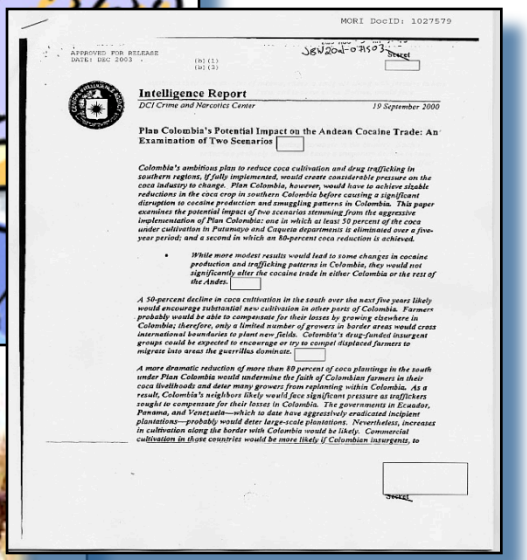
Public Health Preparedness

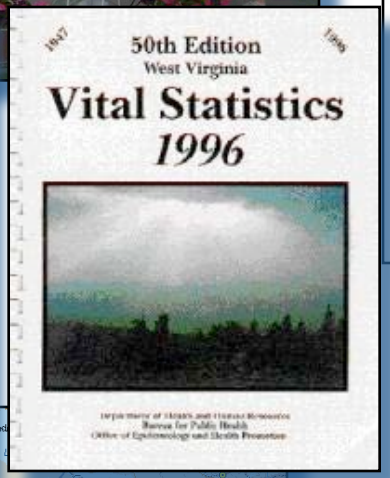
Context is important







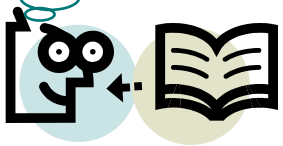




State of the art

State of the art

An elephant with 2 trunks and 5 legs!!!

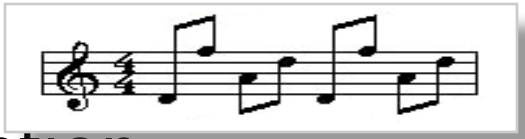


This talk:

- Translational BioInformatics and Information Integration Dilemma
- Case Study: Public Health Preparedness
- Our Vision
- Our Implementation
- Challenges

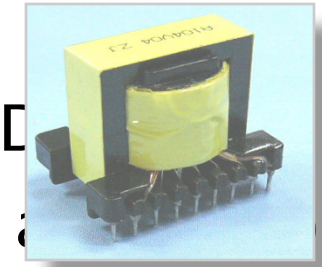
the Solution Framework

- Resource Definition Framework (RDF)



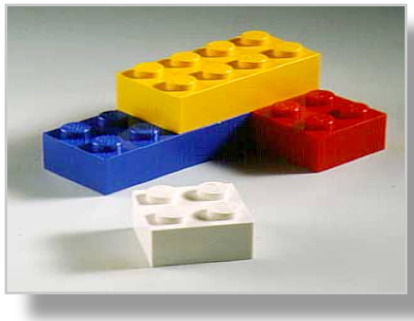
to enable unified Information Representation

- Ontologies (OWL) and Computer Reasoning (D



(OWL-DL) To enable knowledge representation a n

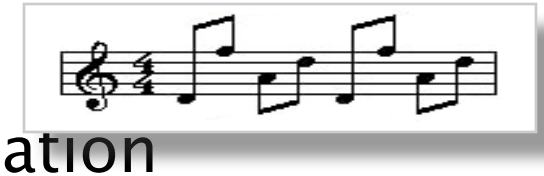
- Services Oriented Architecture



Dynamic interoperability and reuse

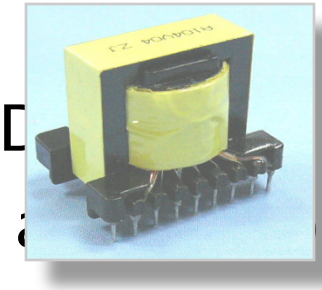
the Solution Framework

- Resource Definition Framework (RDF)



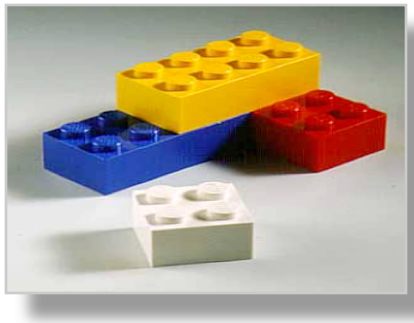
to enable unified Information Representation

Integrative, transdisciplinary, agile, collaborative



- Ontologies (OWL) and Computer Reasoning (D

(OWL-DL) To enable knowledge representation a n

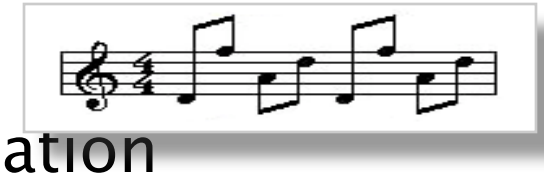


- Services Oriented Architecture

Dynamic interoperability and reuse

the Solution Framework

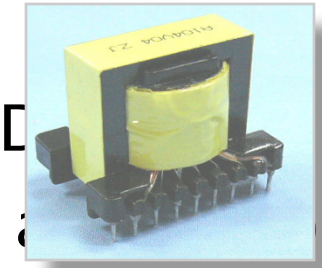
- Resource Definition Framework (RDF)



to enable unified Information Representation

Integrative, transdisciplinary, agile, collaborative

- Ontologies (OWL) and Computer Reasoning (D



(OWL-DL) To enable knowledge representation a n

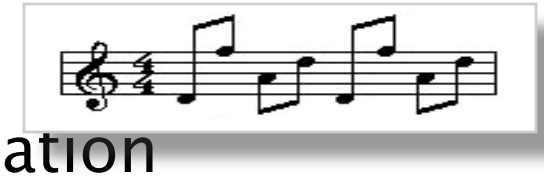
Context aware, knowledge based, agile, transdisciplinary, collaborative

- Services Oriented Architecture

Dynamic interoperability and reuse

the Solution Framework

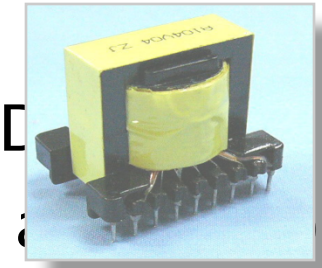
- Resource Definition Framework (RDF)



to enable unified Information Representation

Integrative, transdisciplinary, agile, collaborative

- Ontologies (OWL) and Computer Reasoning (D



(OWL-DL) To enable knowledge representation a n

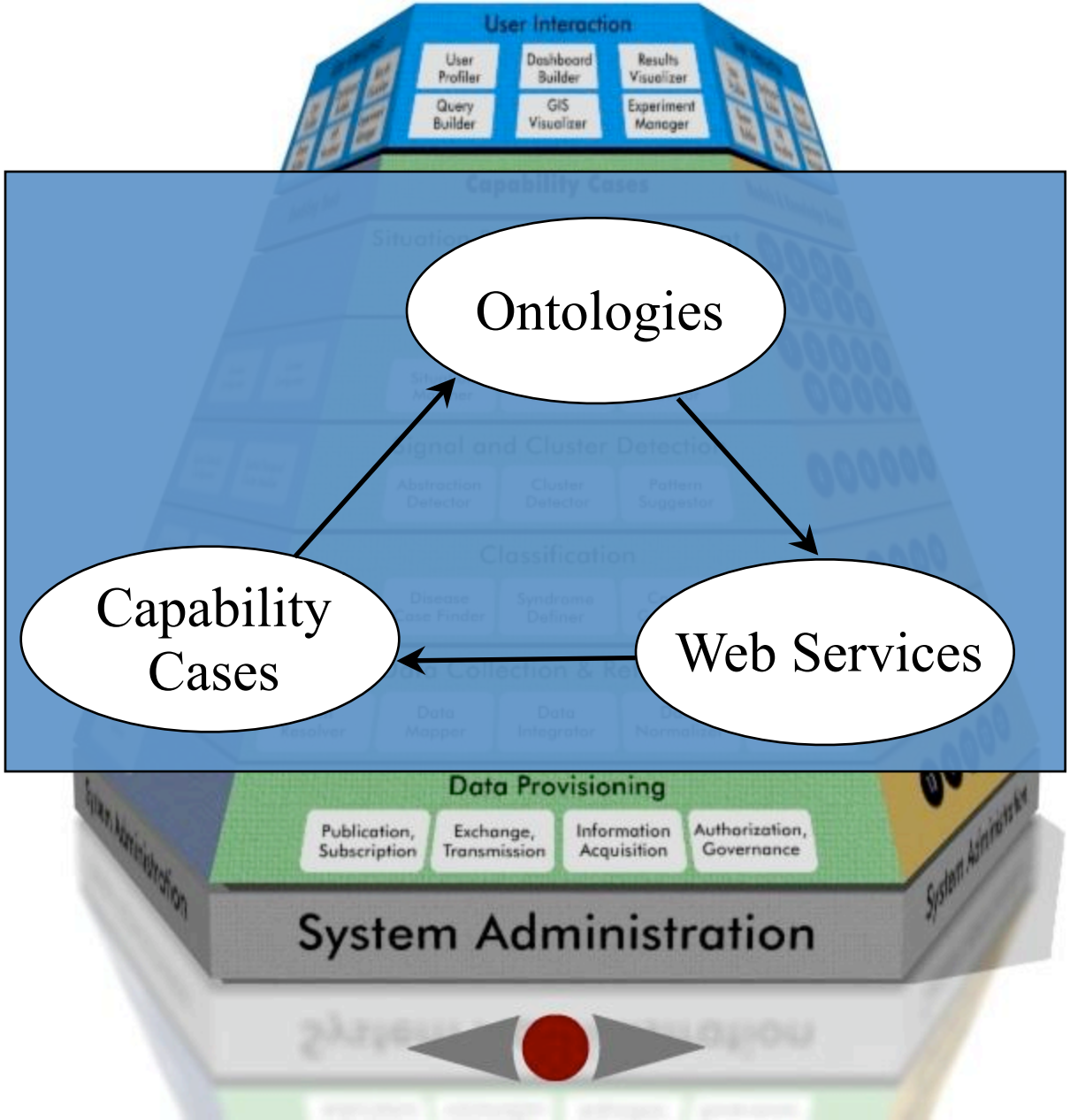
Context aware, knowledge based, agile, transdisciplinary, collaborative

- Services Oriented Architecture

Dynamic interoperability and reuse

Agile, interoperable, collaborative and distributed

Dimensions of SARA



This talk:

- Translational BioInformatics and Information Integration Dilemma
- Case Study: Public Health Preparedness
- Our Vision
- Our Implementation
- Challenges

Data Sources – 1

- Triage Data
  - Patient Demographics (Age, Ethnicity, Gender)
  - Vital Signs (T, RR, PR, PO2)
  - Chief Complaints

Data Sources – 2

- Nurse Notes
  - Vital Signs,
  - Complete Review of Systems: General, Respiratory, Neurological, Gastrointestinal, Dermatological, etc
  - Past Medical and Surgical HX
  - Medications, Past Medications, Home Medications
  - Interventions, Procedures
  - Outcome
  - Discharge and Disposition
  - Past Medical and Surgical HX

Data Sources – 1

- Triage Data
  - Patient Demographics (Age, Ethnicity, Gender)
  - Vital Signs (T, RR, PR, PO2)
  - Chief Complaints
- Nurse Not
  - Vital Si
  - Comple
    Gastroi
  - Past Me
  - Medica
  - Interventions, Procedures
  - Outcome
  - Discharge and Disposition
  - Past Medical and Surgical HX

- From 8 community hospitals and 16 different IT implementations
- Structured, semi-structured, non structured entries
- Automated submissions through HTTP
- Accounts for about %30 Houston ED visits
- Data transmission every 10 minutes or less
- Over 250,000 concepts, 82 million instances and growing

Data Sources - 3

Texas Commission for Environmental Quality (TCEQ)

- Pollution Parameters
  - CO,SO2,H2S,NO, NO2, O3, TNMOC, CH4, ...
- Meteorological Parameters
  - Temperature (Outdoor , Dew Point)
  - Relative Humidity,
  - Radiation (Solar, Ultraviolet, Net Radiation)
  - Barometric Pressure,
  - Precipitation, …
- Chromatography Data
  - Ethane, Methylcyclopentane, 1,2,4-Trimethylbenzene, Ethylene, 2,4-

Data Sources – 3

Texas Commission for Environmental Quality (TCEQ)

- Pollution Parameters
  - …
- Me
  - Temperature (Outdoor , Dew Point)
  - Relative Humidity,
  - Radiation (Solar, Ultraviolet, Net Radiation)
  - Barometric Pressure,
  - Precipitation, …
- Chromatography Data
  - Ethane, Methylcyclopentane, 1,2,4-Trimethylbenzene, Ethylene, 2,4-

- From 18 locations 2 sensors each
- Data Transmission from TCEQ hourly
- 250 concepts on each message
- Air Quality indices calculated twice daily

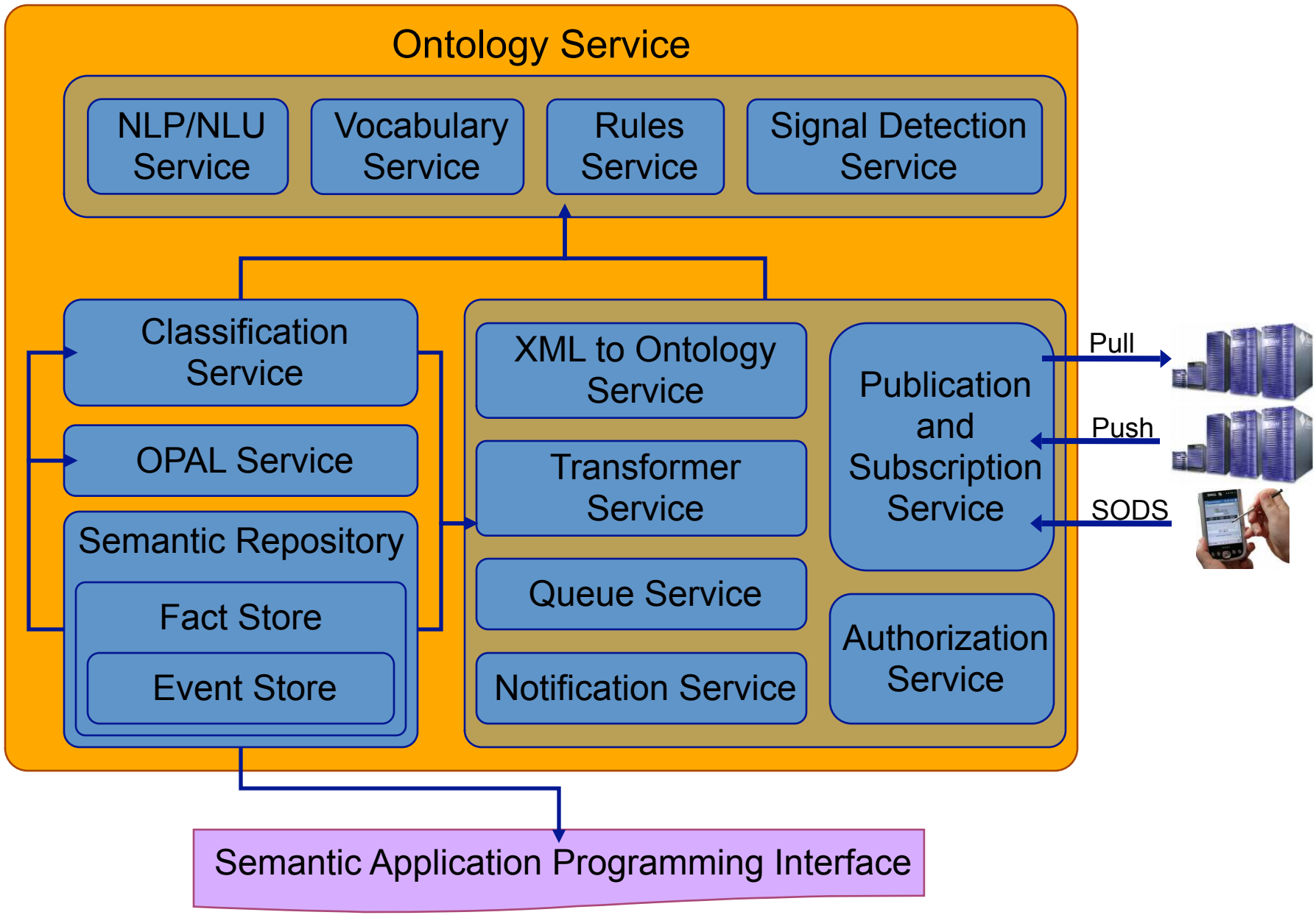
file:///C:/Documents%20and%20Settings/phinformatics/My%20Documents/MHHS/Patient.xml

```
- <Patient VisitId="130996" PersonId="85627293" Gender="Male" Race="Caucasian / White" Age="61 years" InitTemperature="Temp 97.2F Oral" InitBloodPressure="BP 152 / 89"
  InitPulse="Pulse 104bpm" InitPulseOximetry="Pulse Ox 96% Flow RA Room Air">
    <Complaint Value="Other" Custom_Text="Bleeding Toe"/>
  - <Charts>
    - <Chart Type="Nurse">
      - <Section Name="Observations">
        - <Choice Choice_Name="" Choice_Display_Name="Activity" StrikeOut="False">
            <Choice2 Choice_Name="Resting quietly." Choice_Display_Name="Resting quietly" StrikeOut="False"/>
            <Choice2 Choice_Name="No apparent distress." Choice_Display_Name="No apparent distress" StrikeOut="False"/>
          </Choice>
        - <Choice Choice_Name="" Choice_Display_Name="Awaiting" StrikeOut="False">
            <Choice2 Choice_Name="Awaiting disposition." Choice_Display_Name="Disposition" StrikeOut="False"/>
          </Choice>
        </Section>
        <Section Name="Presentation"/>
        <Section Name="Historical"/>
        <Section Name="Screening"/>
        <Section Name="Tri. Assess"/>
      - <Section Name="Vitals">
          <Choice Choice_Name="Vitals" Choice_Display_Name="Vitals" StrikeOut="False" Plugin_Name="VitalSigns" Plugin_Value="BP 158 / 93"/>
          <Choice Choice_Name="Vitals" Choice_Display_Name="Vitals" StrikeOut="False" Plugin_Name="VitalSigns" Plugin_Value="BP 159 / 78"/>
        </Section>
        <Section Name="Assessment"/>
        <Section Name="Interventions"/>
      - <Section Name="Procedures">
          <Choice Choice_Name="No Procedures done.
        </Section>
      - <Section Name="Medication">
        - <Choice Choice_Name="Meds follow up: " Choice_Display_Name="Meds follow up" St                ut="False">
            <Choice2 Choice_Name="No adverse reaction" Choice_Display_Name="No adverse reaction" StrikeOut="False"/>
          </Choice>
        </Section>
      - <Section Name="Outcome">
        - <Choice Choice_Name="" Choice_Display_Name="Condition" StrikeOut="False">
            <Choice2 Choice_Name="Condition: stable" Choice_Display_Name="Stable" StrikeOut="False"/>
            <Choice2 Choice_Name="Condition: good" Choice_Display_Name="Good" StrikeOut="False"/>
          </Choice>
```

Plugin_Value="BP 159 / 78"

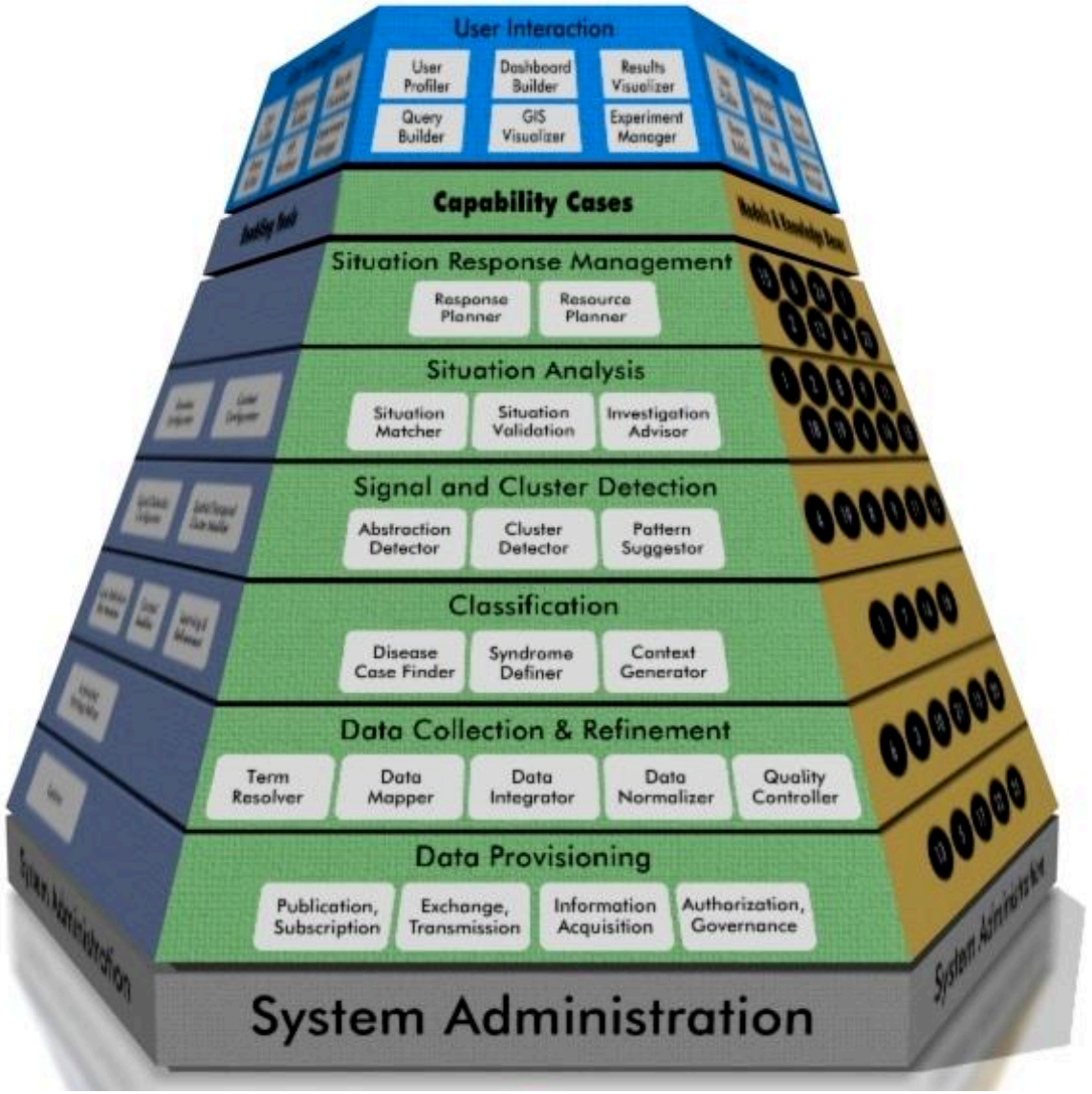
Choice ="No adverse reaction"

Semantic Integration



SAPPHIRE Implementation

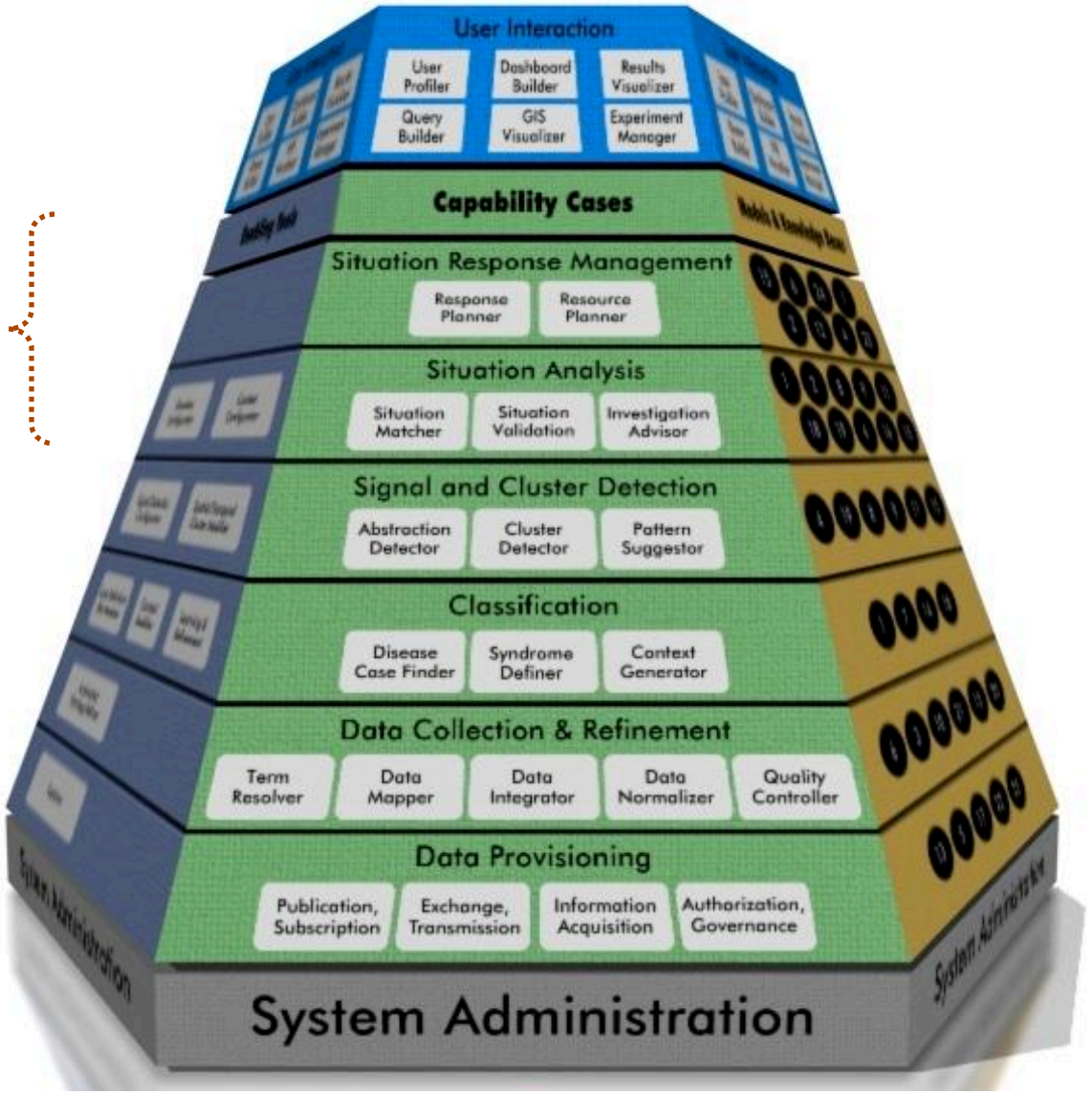
We Are Here



SAPPHIRE Implementation

Proof of concept

We Are Here



Implementation Platform

1- TopBraid Composer as Ontology Management Tool
2- Jena from HP as API for Semantic Web
3- Eclipse Java Development Environment
4- Oracle Semantic Data Management (Started with 10gR2 on Windows, Currently 11gR1 on Linux)
5- Pellet/Jena OWL Micro Reasoner
6- Services Oriented Architecture
7- Microsoft SQL Server 2005 XML archive and Analysis Services
8- IBM Dual Xeon 2.8GH/3GB RAM Blade Server
9- EqualLogic iSCSI SAN (4 TB)
10- GB Ethernet LAN

This talk:

- Translational BioInformatics and Information Integration Dilemma
- Case Study: Public Health Preparedness
- Our Vision
- Our Implementation
- Challenges

Challenges

- State of the frameworks
- Maturity of Tools
- Knowledge Engineering and Ontology Development
- Reasoning and Rules Support
- Scalability
- Performance

Academic and Industrial

- Scalable, High Performance RDF/OWL Repositories (Oracle, Franz)
- Scalable Semantic Application Programming Interface (Oracle, TopQuadrant, HP)
- Ontology based Business Intelligence and Data Mining (TopQuadrant)

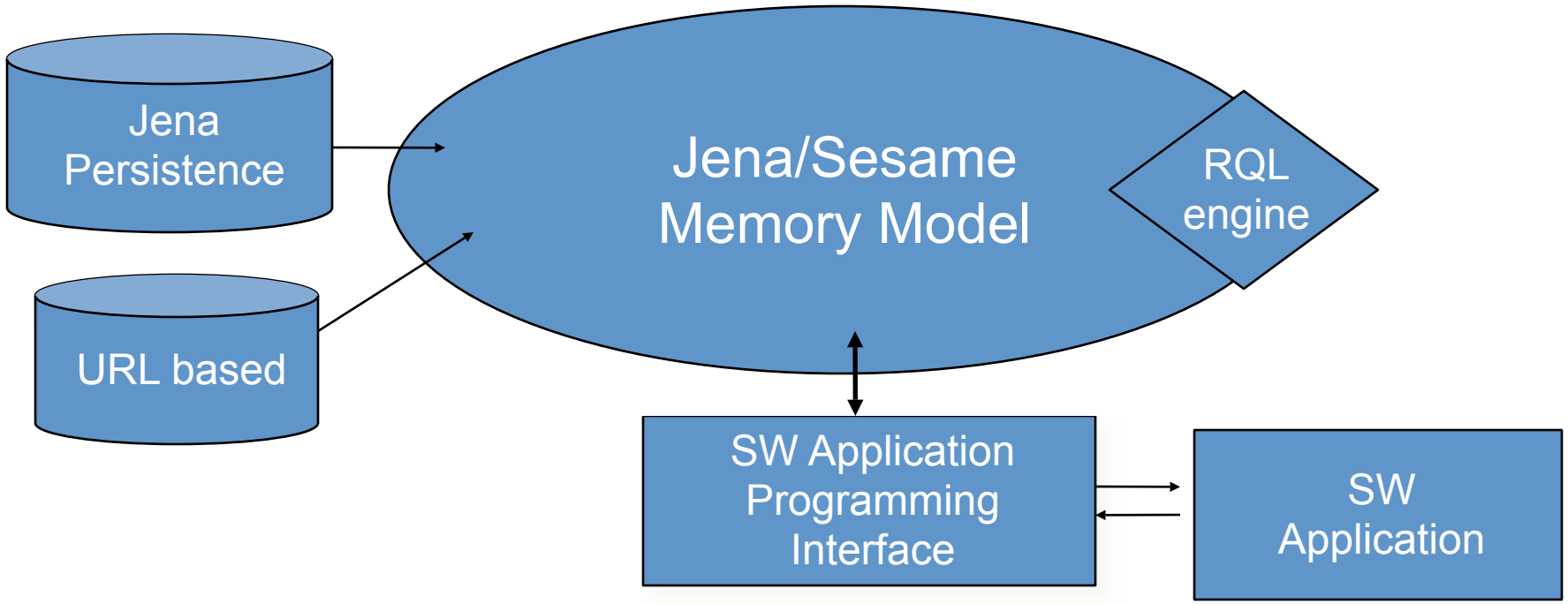
Scalable Semantic Application Programming

- Scalable SW application development interface for Oracle Semantic Data Management (SDM)
- Seamless integration of application development interfaces to Oracle SDM
- Without any intermediate or ‘in memory’ representation of semantic data

Method (conceptual)

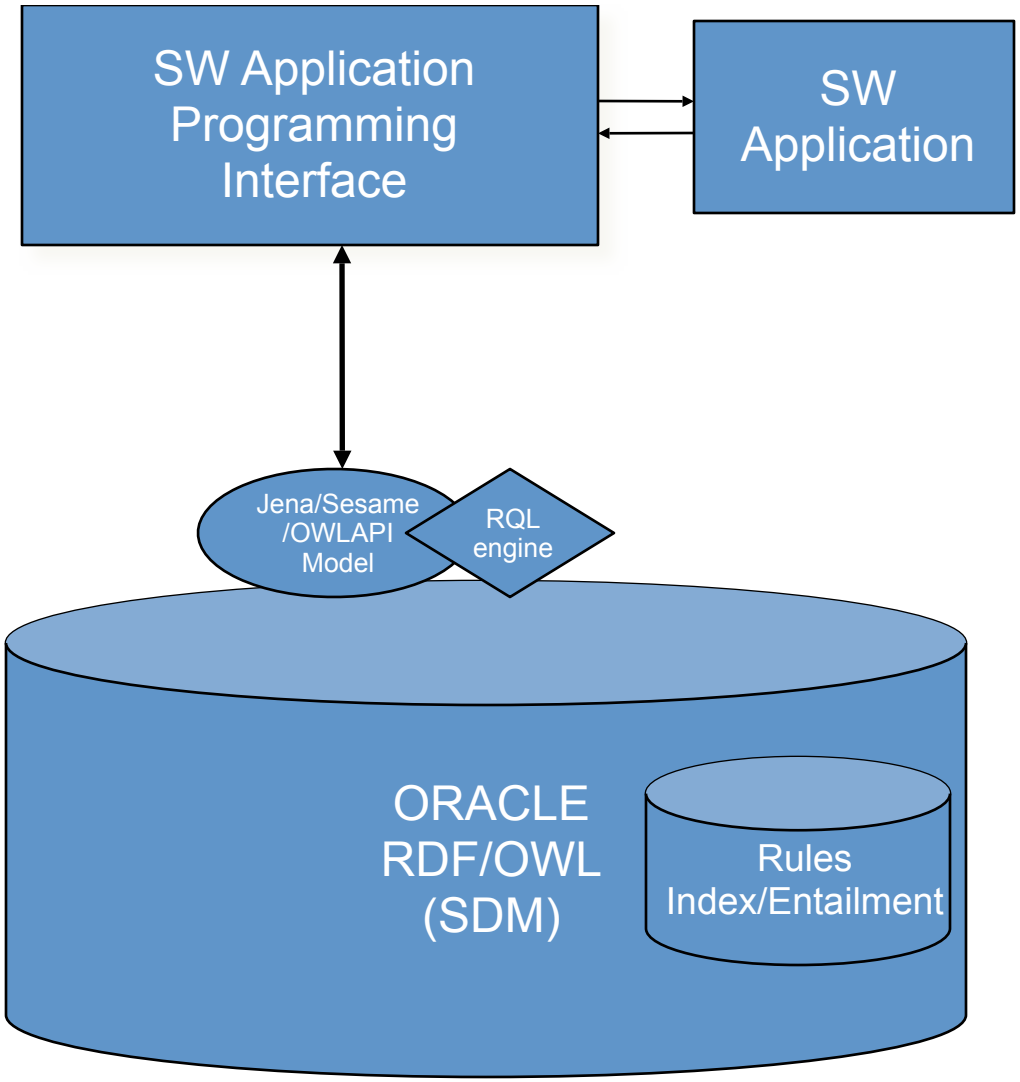
- Adopt existing SW application development frameworks (e.g. Jena)
- Extend by eradicating the intermediate representation of Semantic data in memory
- Support SPARQL in the applications side
- Enable invocation and use of integrated rules engines through API

Jena Conceptual Model As Is

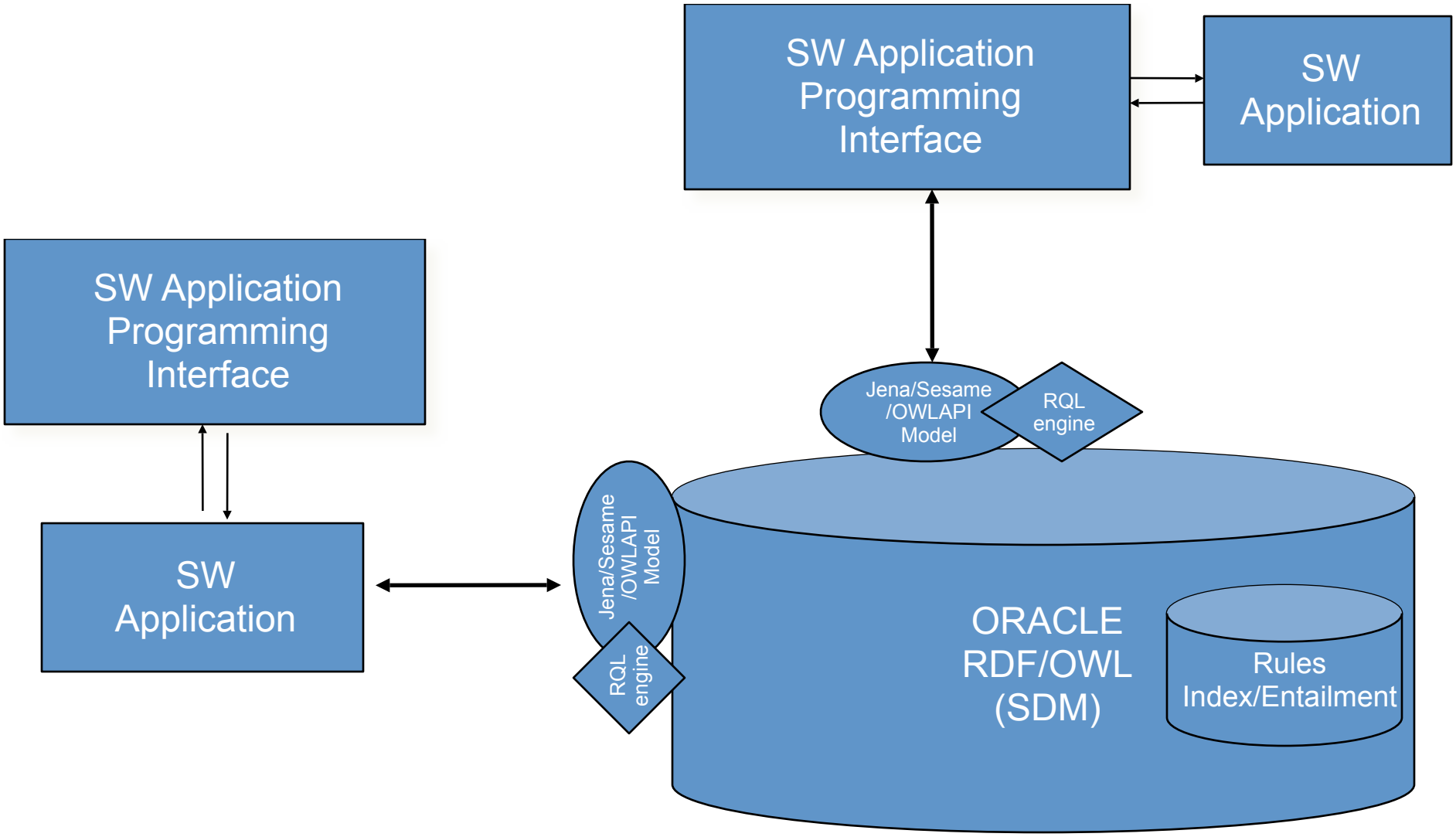


- Not scalable, depends on memory resources
- Does not support concurrent users and distributed applications
- Reasoning is not scalable and integrated with repositories

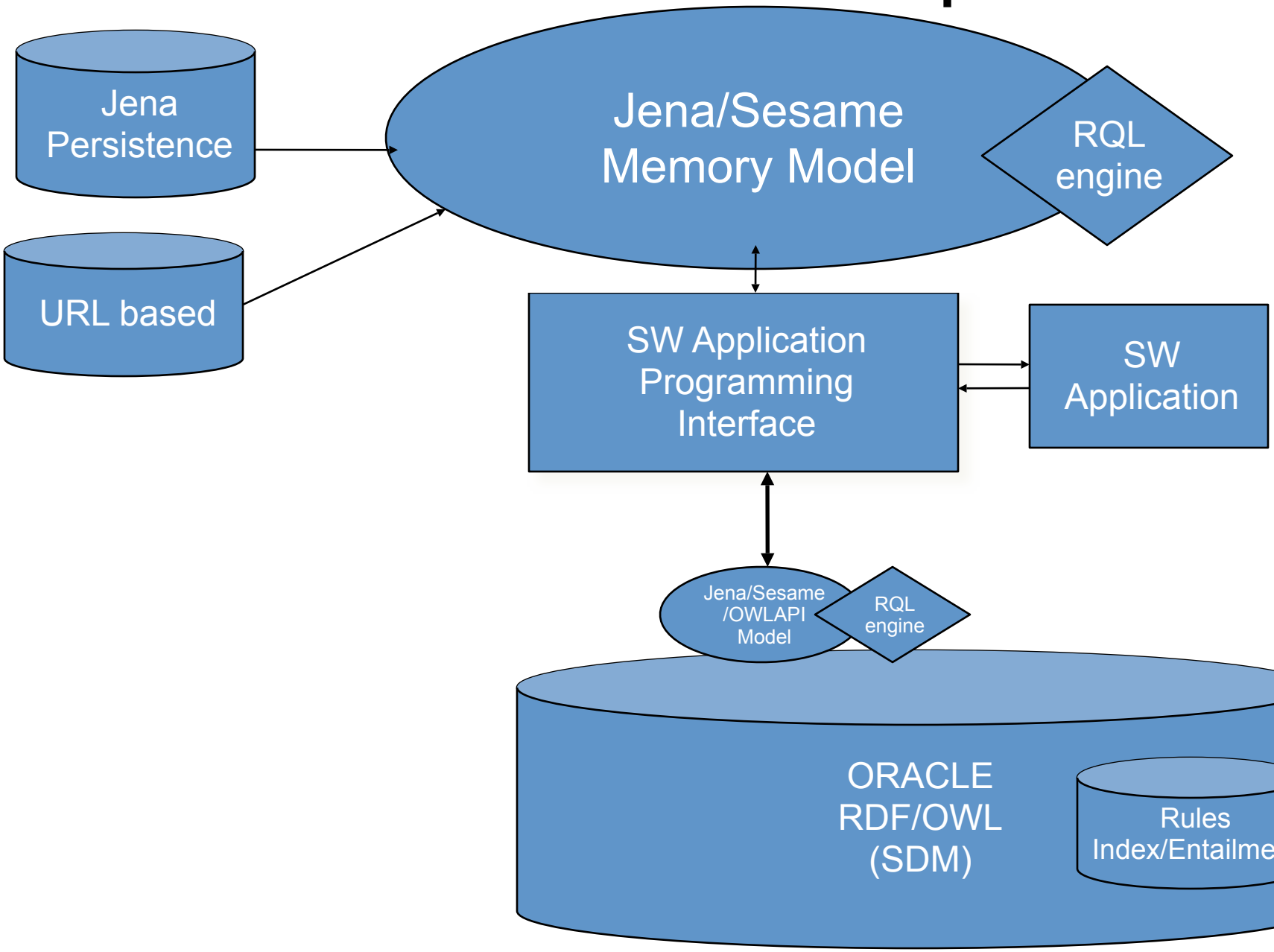
Oracle Jena Adapter



Oracle Jena Adapter



Oracle Jena Adapter



Pros and Cons

- Enables use of Oracle SDM for large scale implementations using Graph and Model objects
- Complete but indirect support of SPARQL
- Supports multi-user and distributed application environments
- Integrated Support of Oracle Reasoners (RDFS and OWLPrime) in the application side
- Robust performance through both programming interface as well as SPARQL querying

- Indirect support of SPARQL (not available outside of the API through SQL Developer for example)
- OntModel to be supported in future releases
- Rules reasoning not real time

Contact

- Parsa Mirhaji, MD Director

  The Center for Biosecurity and Public Health Informatics Research The School of Health Information Sciences The University of Texas - Health Sciences Center at Houston Office: (713) 500-3157 Fax: (713) 500-0370 Assistance (Namiko Burleson): (713) 500-3938

- http://www.phinformatics.org/ResearchProjects/SAPPHIRE/tabid/76/Default.aspx