8-14-2007

Sensor Data Management

Cory Andrew Henson

Wright State University - Main Campus

KNO.E.SIS

COLLECTING THE DOTS | CONNECTING THE DOTS

Sensor Data Management

Presentation Outline

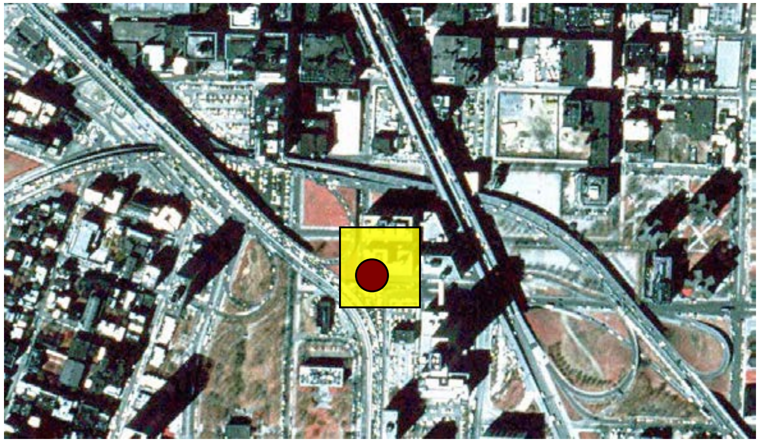
1. Motivating Scenario
2. Sensor Web Enablement
3. Sensor data evolution hierarchy
4. Semantic Analysis

Motivating Scenario

High-level Sensor (S-H)

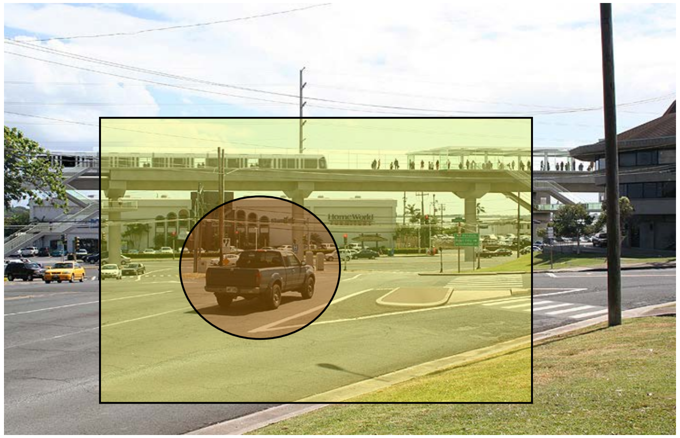
Low-level Sensor (S-L)

H



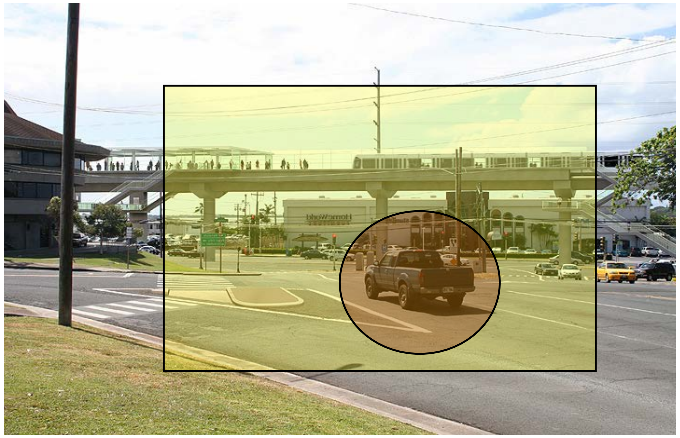
A-H E-H

L



A-L E-L

- How do we determine if A-H = A-L? (Same time? Same place?)
- How do we determine if E-H = E-L? (Same entity?)
- How do we determine if E-H or E-L constitutes a threat?



The Challenge

Collection and analysis of information from heterogeneous multi-layer sensor nodes

Why is this a Challenge?

- There is a lack of uniform operations and standard representation for sensor data.
- There exists no means for resource reallocation and resource sharing.
- Deployment and usage of resources is usually tightly coupled with the specific location, application, and devices employed.
- **Resulting in a lack of interoperability.**

The Solution

The Open Geospatial Consortium Sensor Web Enablement Framework

Presentation Outline

1. Motivating Scenario
2. Sensor Web Enablement
3. Sensor data evolution hierarchy
4. Semantic Analysis

Open Geospatial Consortium

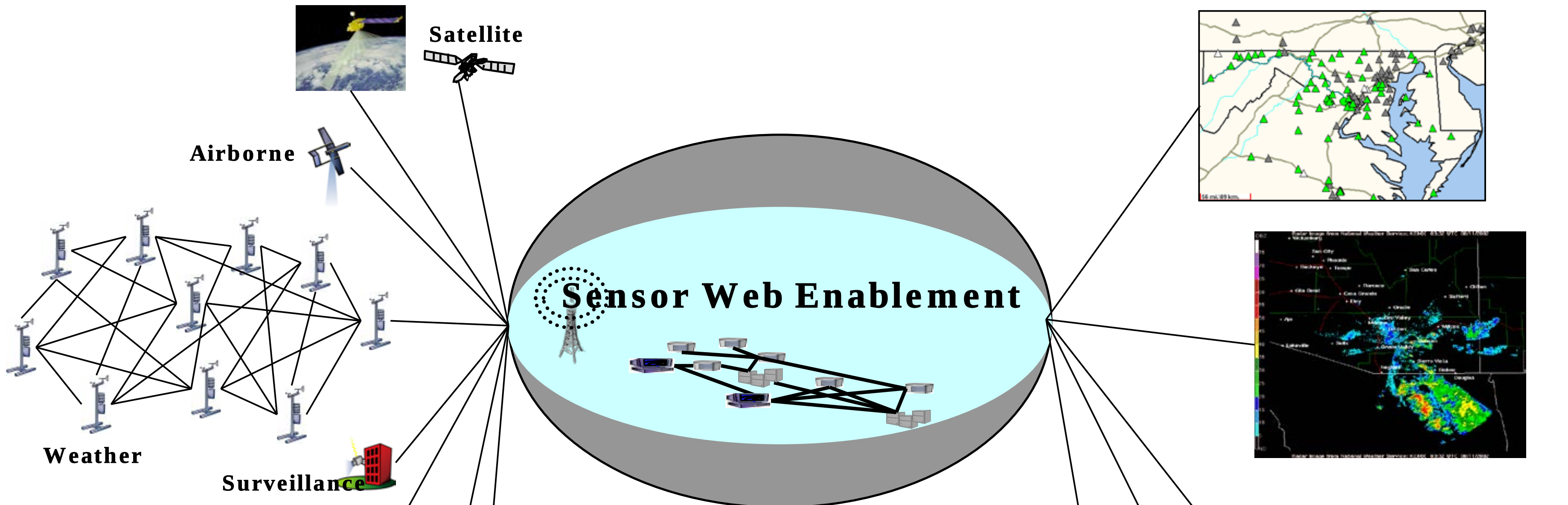
- Consortium of 330+ companies, government agencies, and academic institutes
- Open Standards development by consensus process
- Interoperability Programs provide end-to-end implementation and testing before spec approval
- **Standard encodings**, e.g.
  - GeographyML, SensorML, Observations & Measurements, TransducerML, etc.
- Standard Web Service interfaces, e.g.
  - Web Map Service
  - Web Feature Service
  - Web Coverage Service
  - Catalog Service
  - **Sensor Web Enablement Services** (Sensor Observation Service, Sensor Alert Service, Sensor Process Service, etc.)

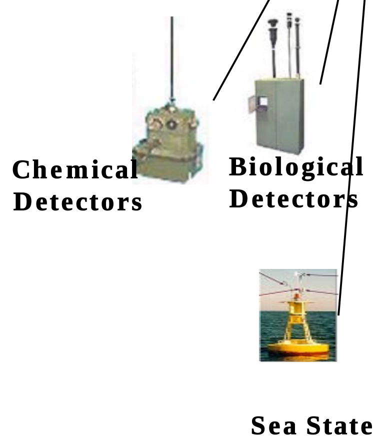
OGC Mission

To lead in the development, promotion and harmonization of open spatial standards

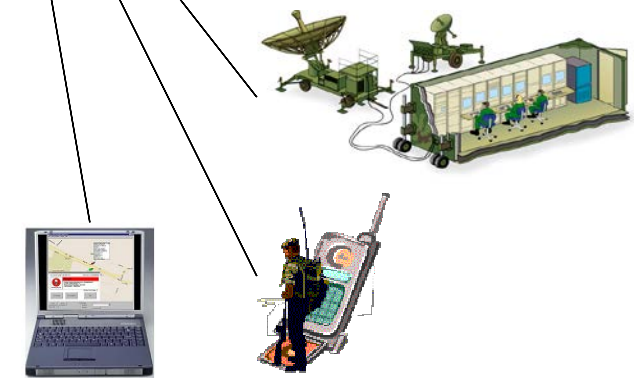
Sensor Web Enablement

Constellations of heterogeneous sensors — **Vast set of users and applications**

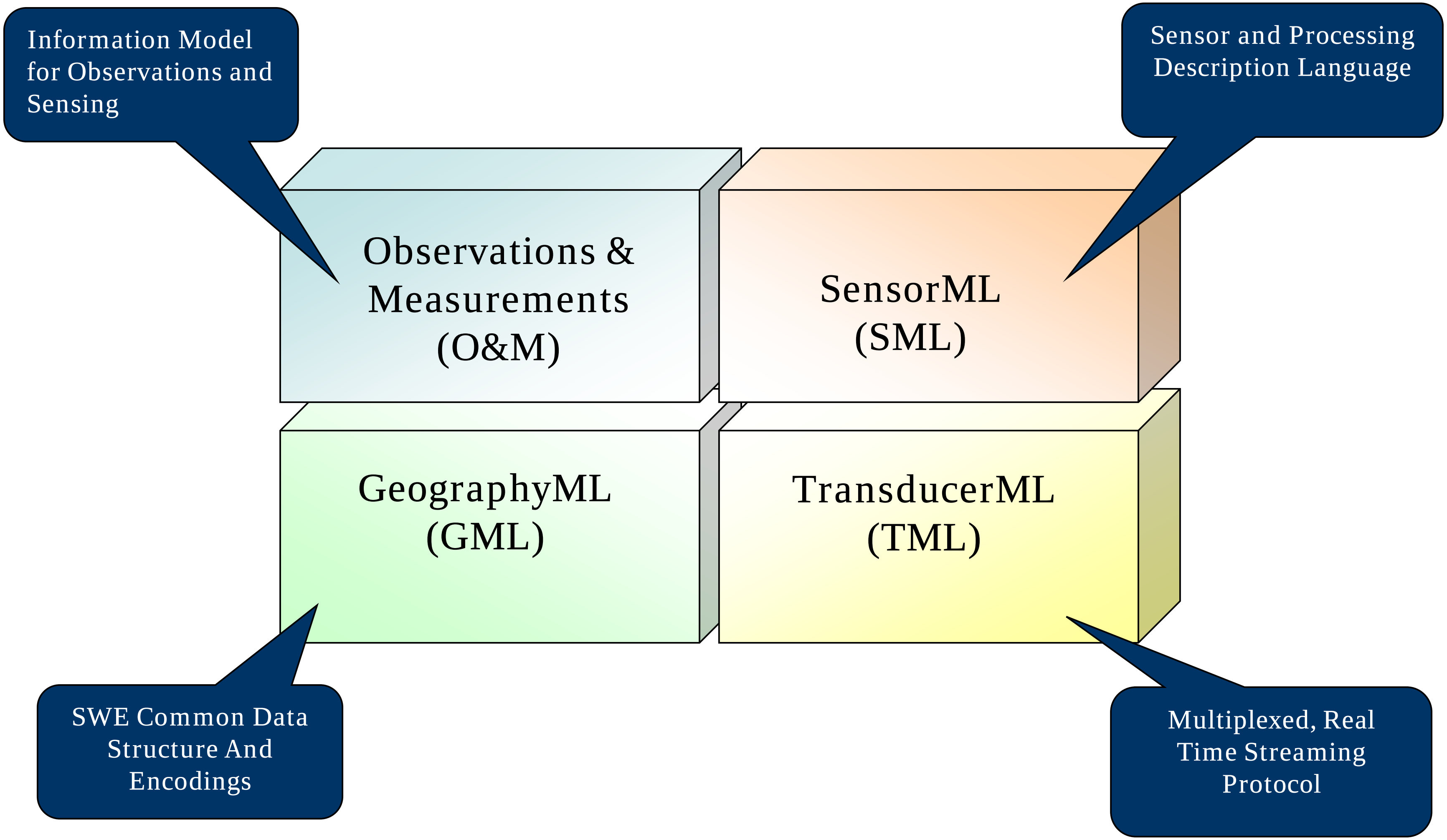




- Distributed self-describing sensors and related services
- Link sensors to network and network-centric services
- Common XML encodings, information models, and metadata for sensors and observations
- Access observation data for value added processing and decision support applications
- Users on exploitation workstations, web browsers, and mobile devices

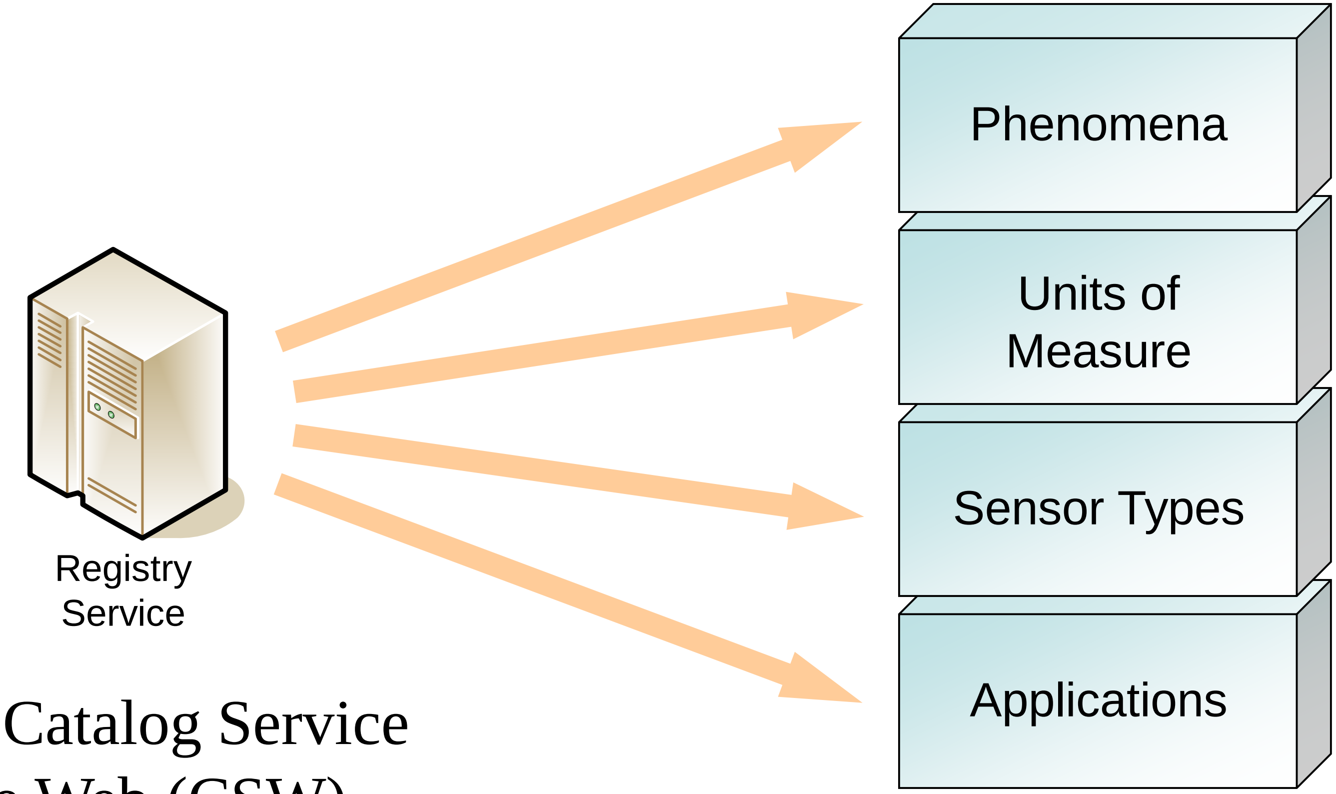


SWE Languages and Encodings



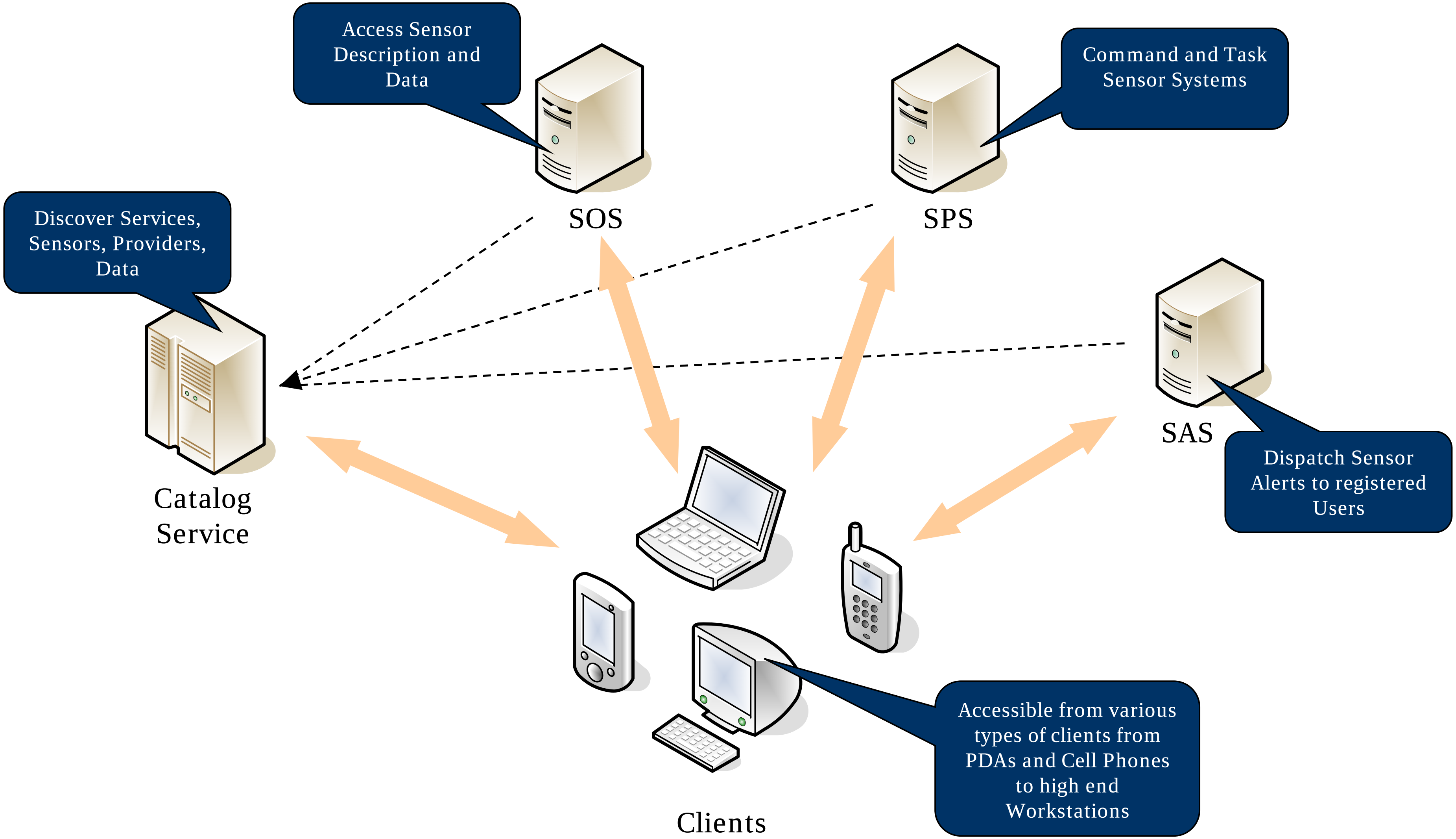
Sam Bacharach, "GML by OGC to AIXM 5 UGM," OGC, Feb. 27, 2007.

SWE Components - Dictionaries



OGC Catalog Service
for the Web (CSW)

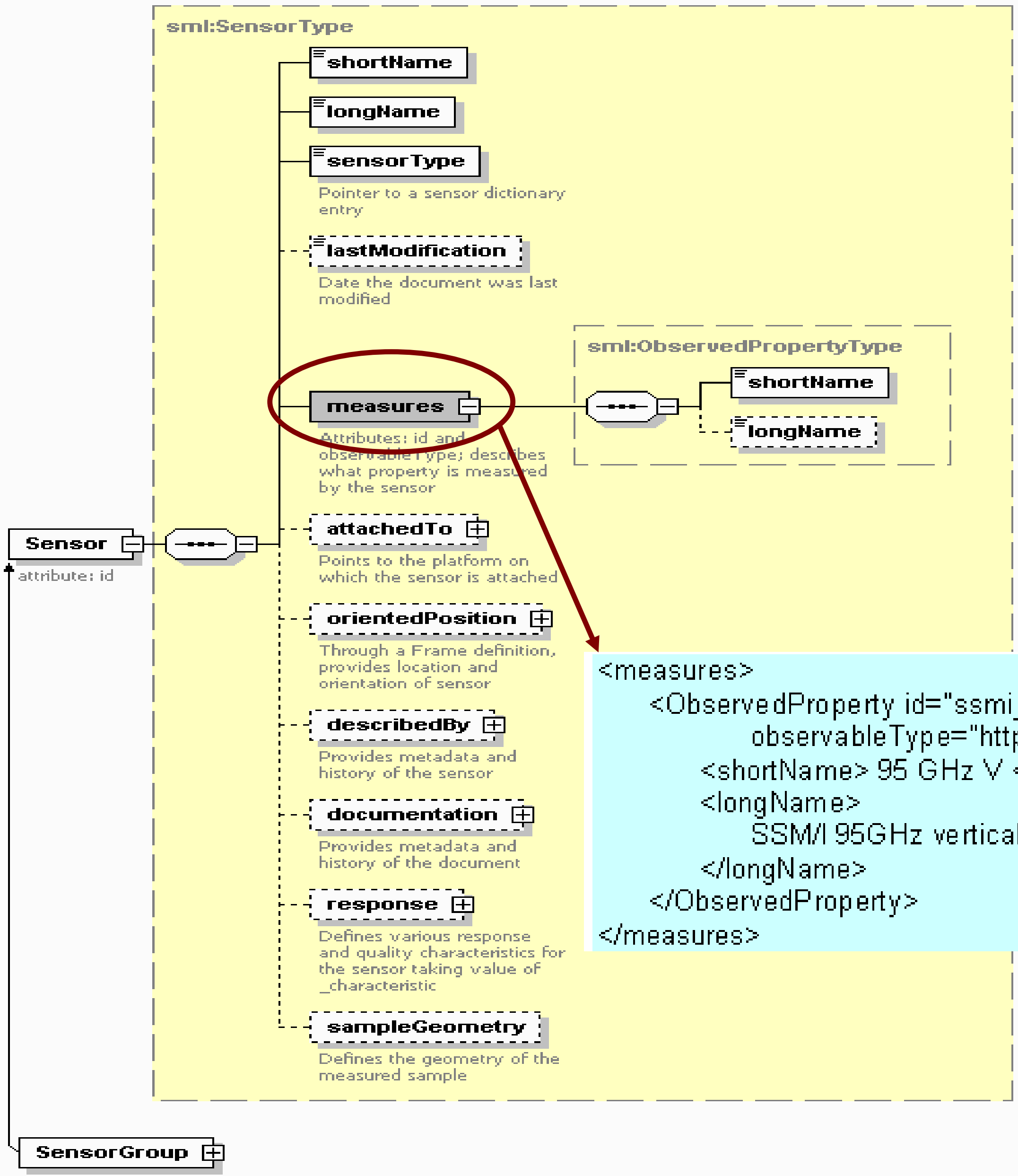
SWE Components – Web Services



Sam Bacharach, “GML by OGC to AIXM 5 UGM,” OGC, Feb. 27, 2007.

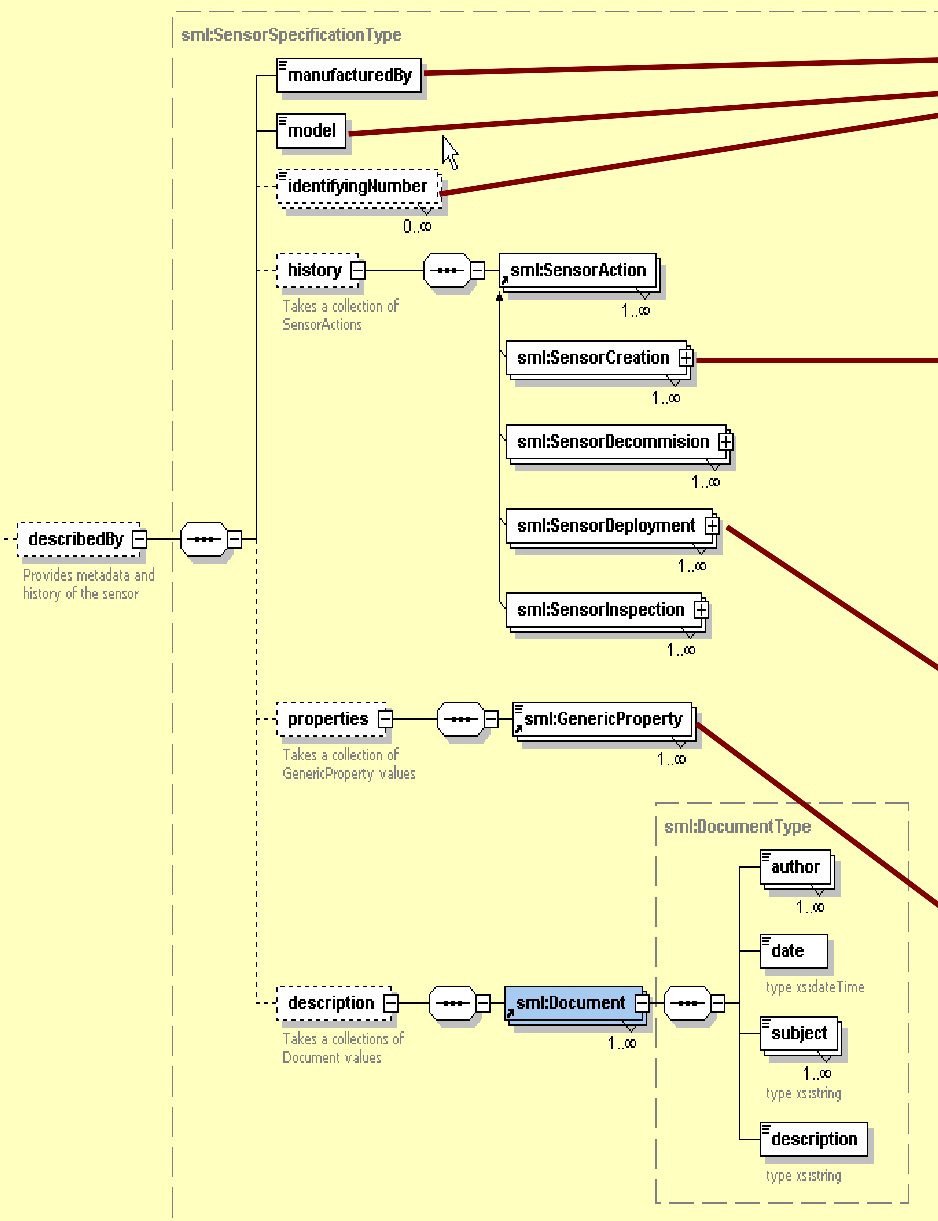
Sensor Model Language
(SensorML)

SML Concepts – Sensor



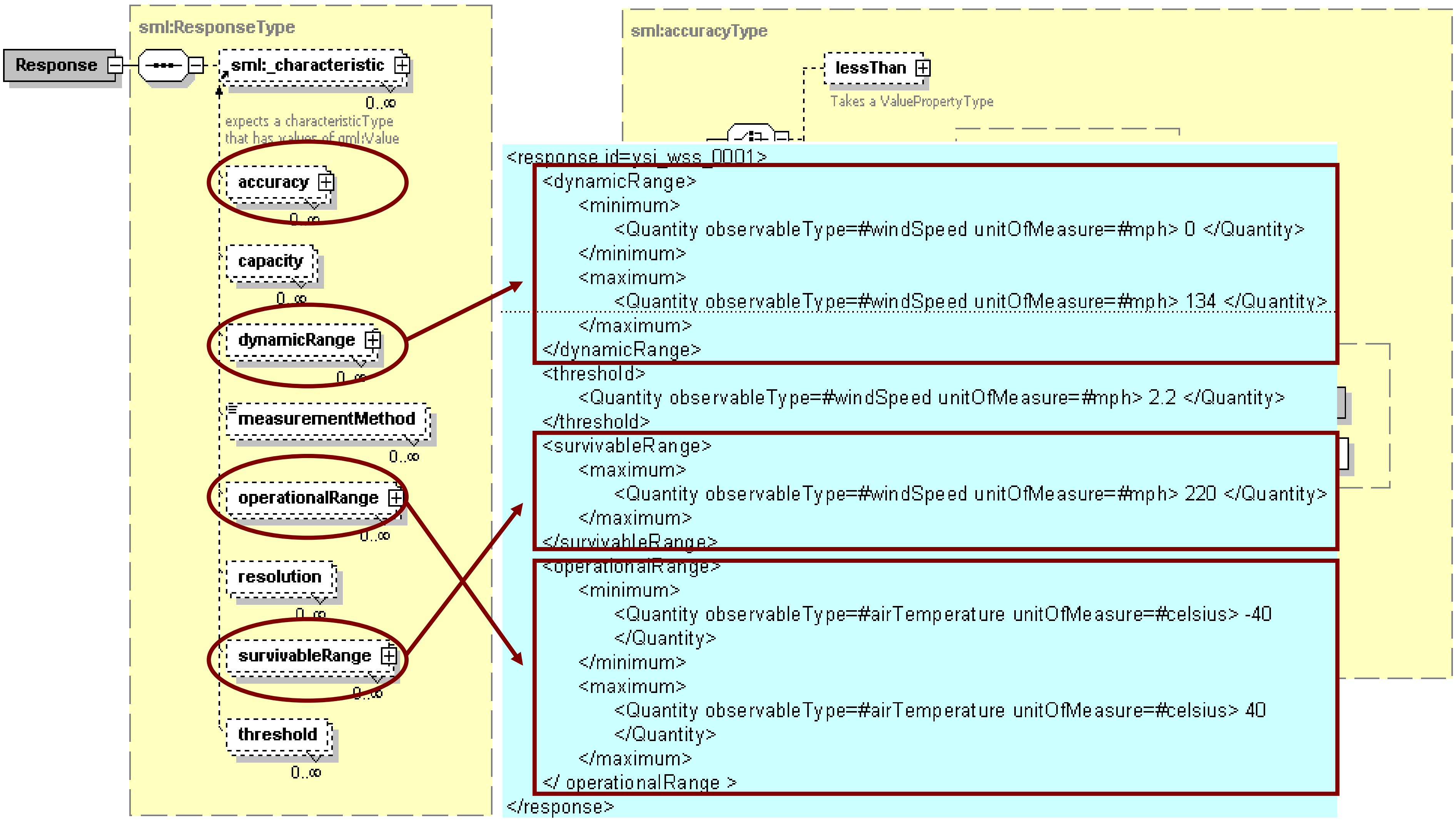
```
<measures>
    <ObservedProperty id="ssmi_f10_95V"
            observableType="http:www.opengis.net/observation/type#radiation">
        <shortName> 95 GHz V </shortName>
        <longName>
            SSM/I 95GHz vertical polarization channel on DMSP F10 satellite
        </longName>
    </ObservedProperty>
</measures>
```

SML Concepts – Sensor Description



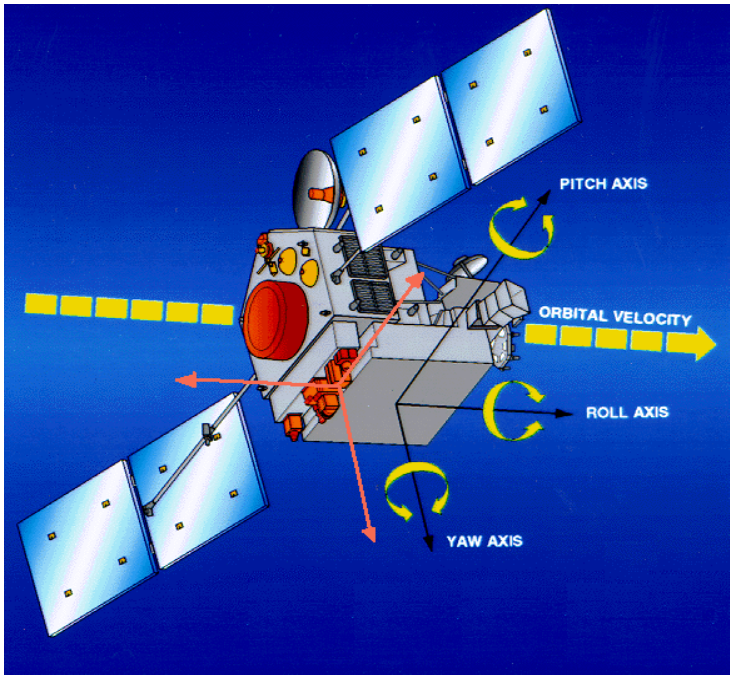

```
<describedBy>
    <SensorSpecification>
        <manufacturedBy> YSI </manufacturedBy>
        <model> 58 </model>
        <identifyingNumber type="serialNumber"> s454165 </identifyingNumber>
        <history>
            <SensorCreation>
                <byWhom>
                    <Person>
                        <fullName> Some Guy </fullName>
                    </Person>
                </byWhom>
                <when> 2001-12-04 </when>
                <supportingDocuments>
                    <Reference id="YSI-12">
                        <authors>
                            <Person>
                                <fullName> Some Author </fullName>
                            </Person>
                        </authors>
                        <publishingDate> 2001-01-01 </publishingDate>
                        <documentTitle> Blueprints for This Sensor </documentTitle>
                        <documentNumber> YSI-12345-678 </documentNumber>
                    </Reference>
                </supportingDocuments>
            </SensorCreation>
            <SensorDeployment>
                <byWhom>
                    <Person>
                        <fullName>SomeOther Guy</fullName>
                    </Person>
                </byWhom>
                <where> Someplace, Somewhere </where>
                <when> 2001-12-04 </when>
                <description> Attached to northern most footing on Some Bridge at a depth
                    of 5 meters </description>
            </SensorDeployment>
        </history>
        <properties>
            <genericProperty name="sensorTechnology" dataType="xs:string"> rapid
                pulse </genericProperty>
            <genericProperty name="measurementMethod" dataType="xs:string"> EPA
                accepted </genericProperty>
            <genericProperty name="membraneThickness" dataType="xs:double"
                uom="#mil"> 1.0 </genericProperty>
            <genericProperty name="membraneType" dataType="xs:string"> Teflon
                </genericProperty>
            <genericProperty name="probeSolutionType" dataType="xs:string"> Na2SO4
                </genericProperty>
        </properties>
    </SensorSpecification>
</describedBy>
```

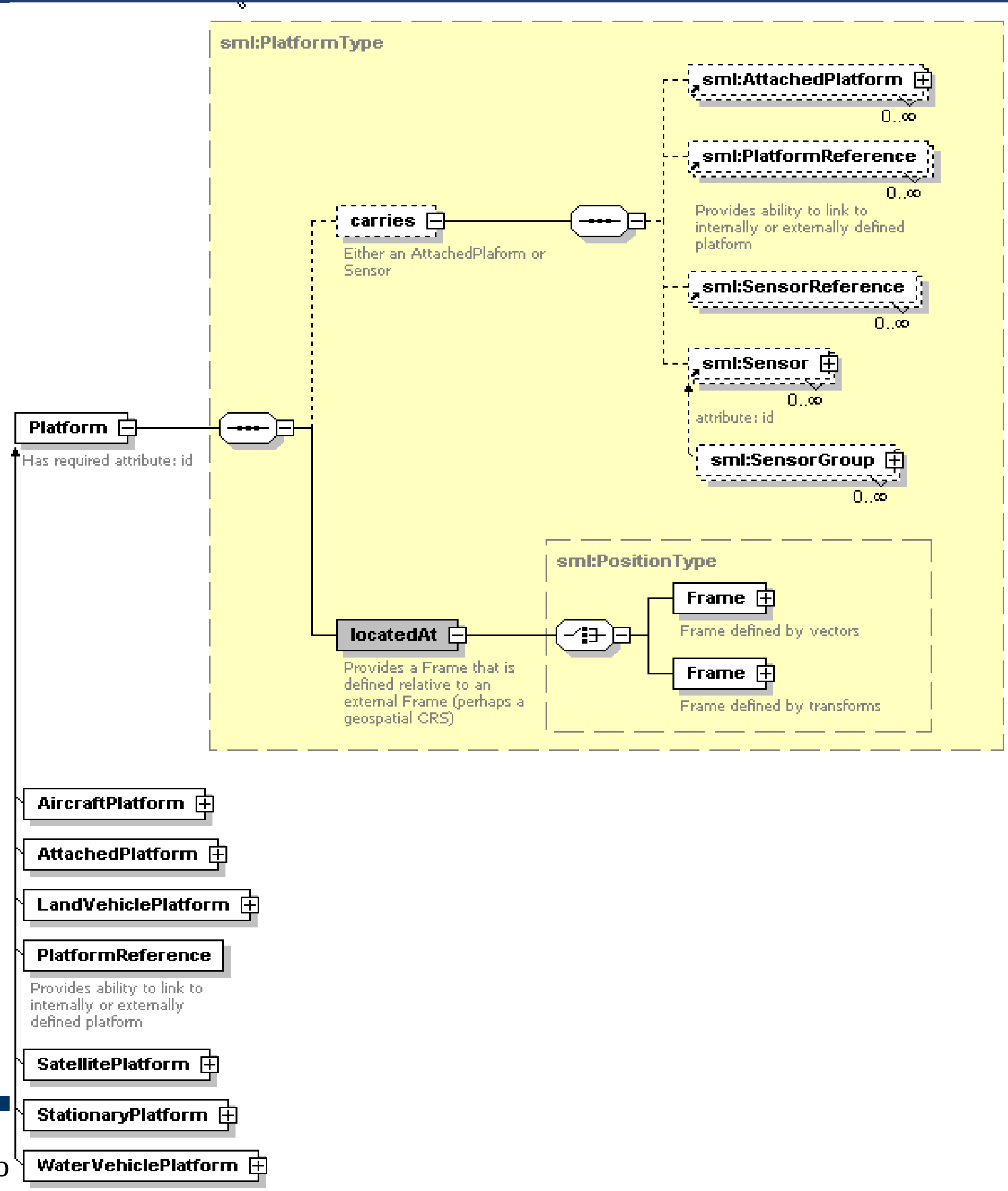
SML Concepts –Accuracy and Range



Mike Botts, "SensorML and Sensor Web Enablement," Earth System Science Center, UAB Huntsville

SML Concepts –Platform

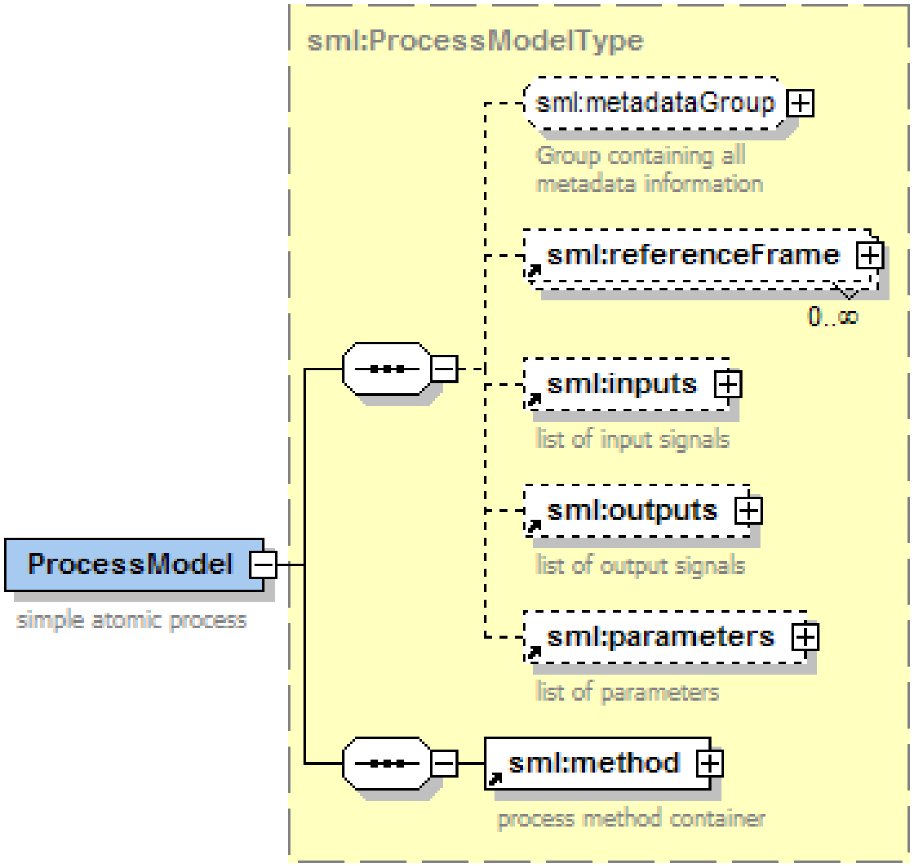




SML Concepts – Process Model

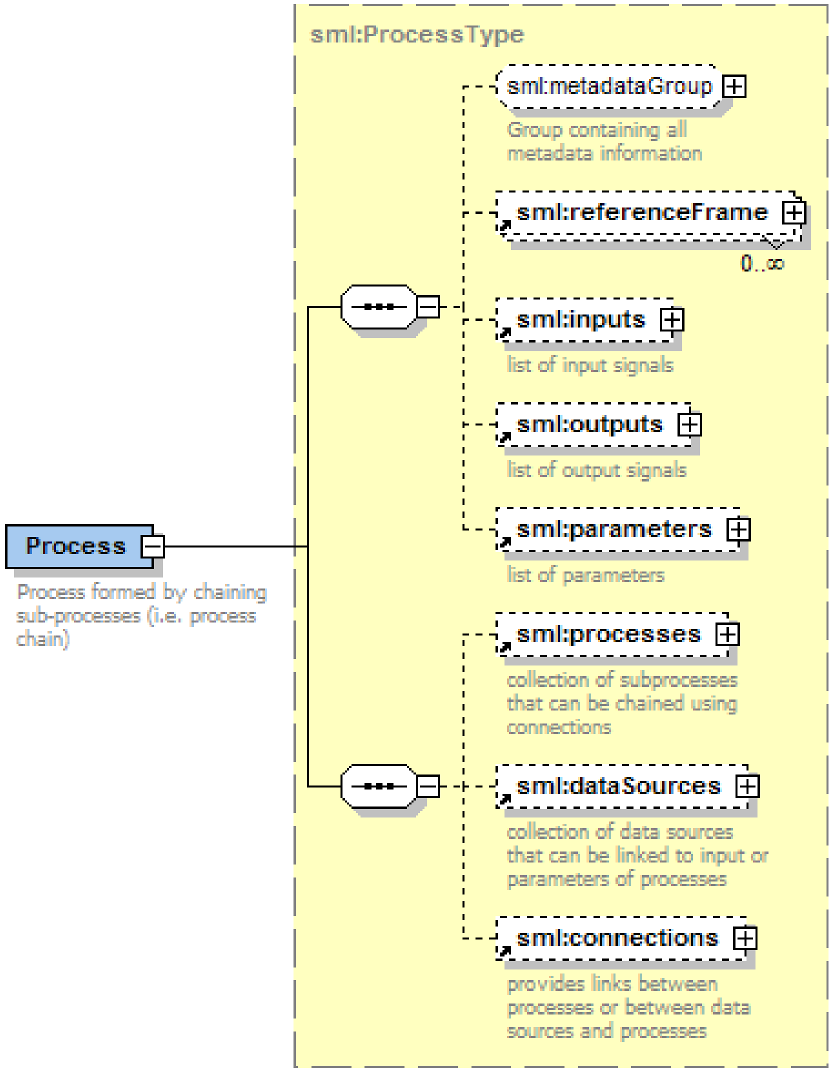
- *In SensorML, everything is modeled as a Process*

- ProcessModel
  - defines atomic process modules (detector being one)
  - has five sections
    - metadata
    - inputs, outputs, parameters
    - method
  - Inputs, outputs, and parameters defined using SWE Common data definitions



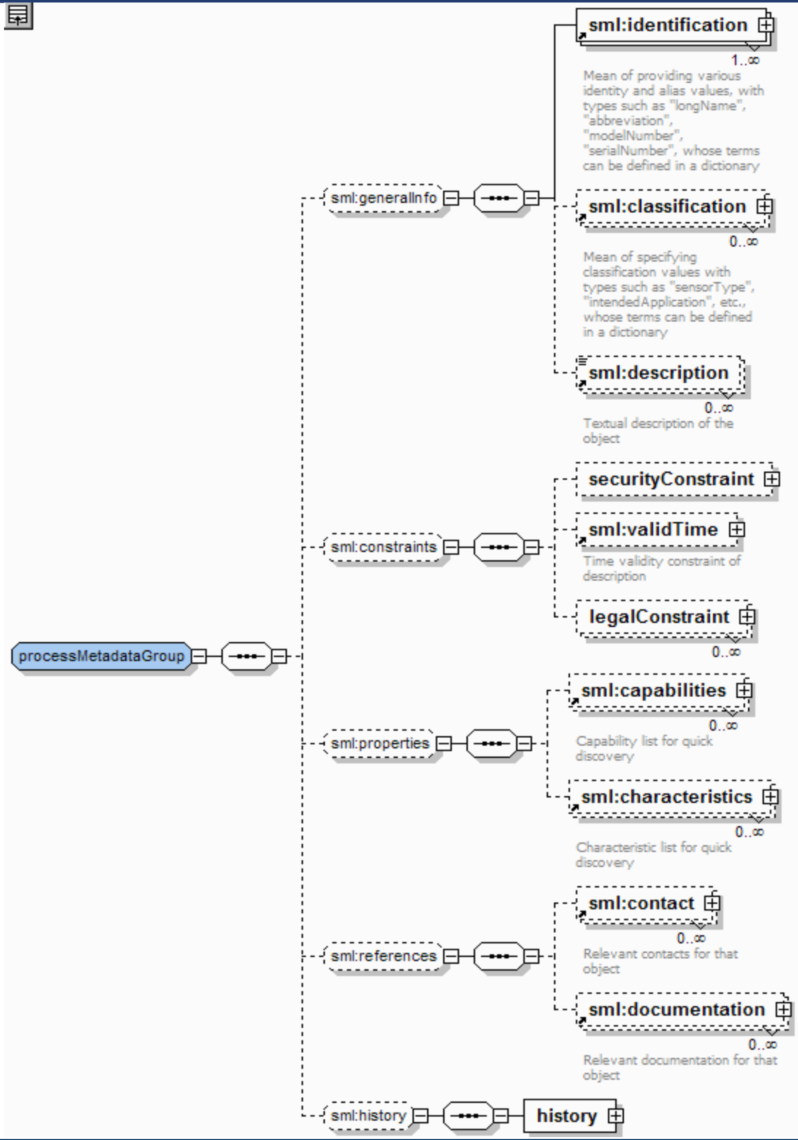
SML Concepts – Process

- <u>Process</u>
  - defines a process chain
  - includes:
    - metadata
    - inputs, outputs, and parameters
    - processes (ProcessModel, Process)
    - data sources
    - connections between processes and between processes and data
- <u>System</u>
  - defines a collection of related processes along with positional information



SML Concepts –Metadata Group

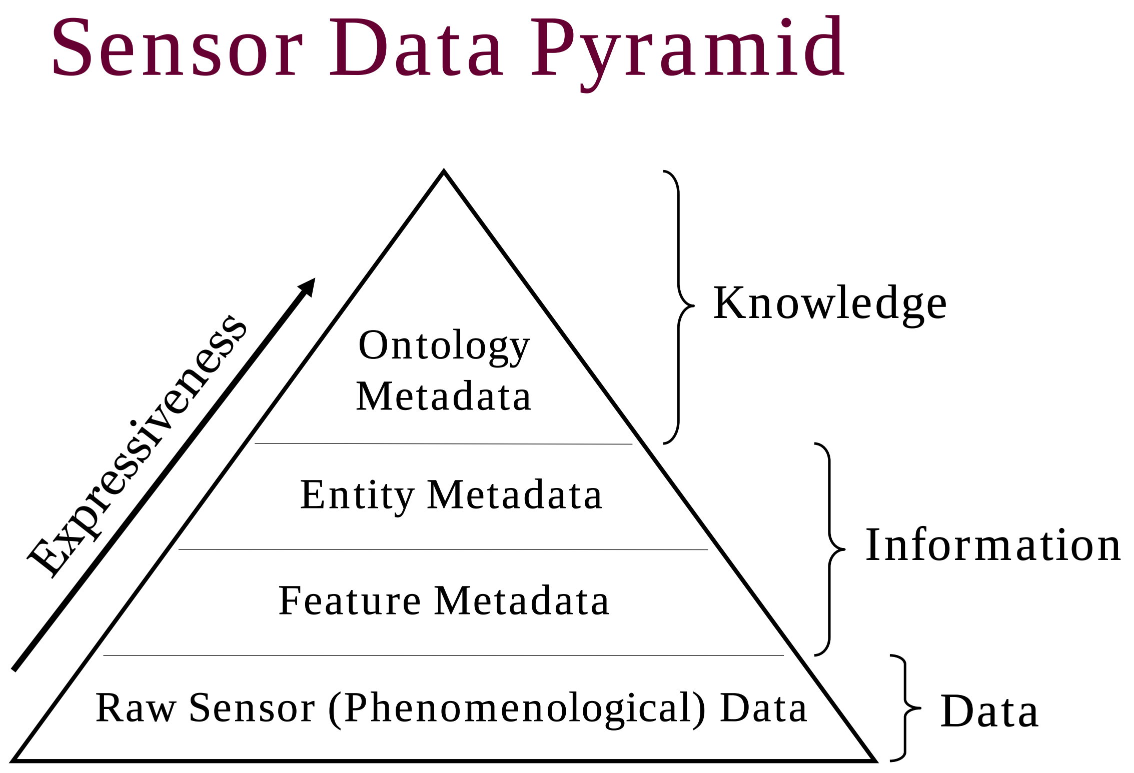
- *Metadata is primarily for discovery and assistance, and not typically used within process execution*
- Includes
  - Identification, classification, description
  - Security, legal, and time constraints
  - Capabilities and characteristics
  - Contacts and documentation
  - History



Presentation Outline

1. Motivating Scenario
2. Sensor Web Enablement
3. <u>Sensor data evolution hierarchy</u>
4. Semantic Analysis

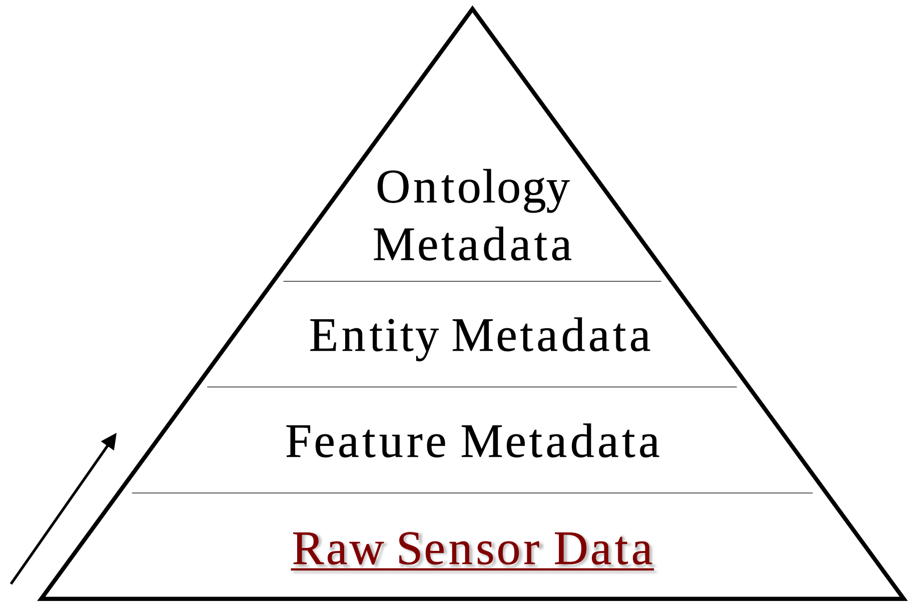
Data Pyramid


Sensor Data Pyramid
Expressiveness
Ontology
Metadata
Entity Metadata
Feature Metadata
Raw Sensor (Phenomenological) Data
Knowledge
Information
Data

Sensor Data Pyramid

Challenges

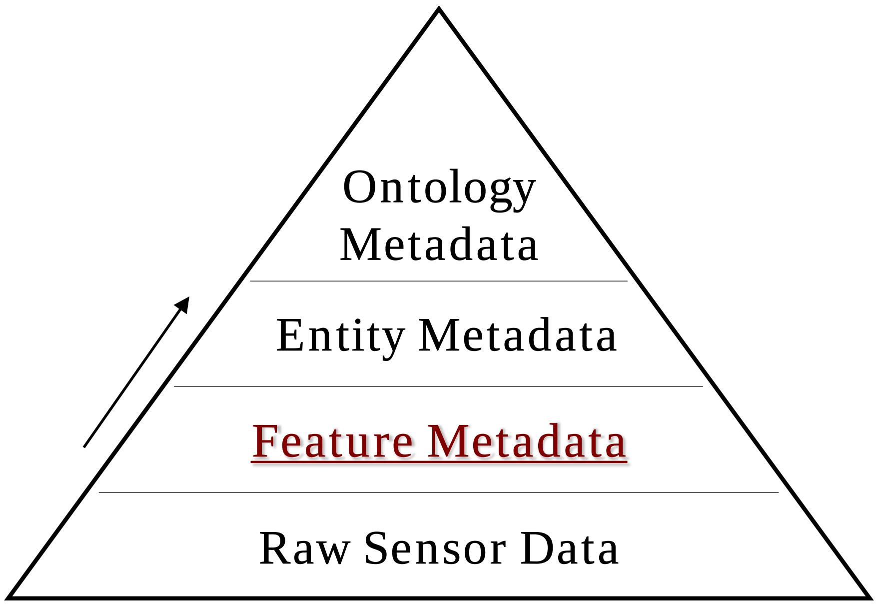
- Avalanche of data
- Streaming data
- Multi-modal/level data fusion
- Lack of interoperability



Solution Goal

1. Collect data from network of multi-level, multi-modal, heterogeneous sensors
2. Annotate streaming sensor data with TransducerML and utilize metadata to enable data fusion
3. Use SensorML to model sensor infrastructure and data processes
4. Annotate sensor data with SensorML
5. Store sensor metadata in XML database
6. Query sensor metadata with XQuery

Sensor Data Pyramid



Challenges

- Extract features from data
- Annotate data with features
- Store and query feature metadata

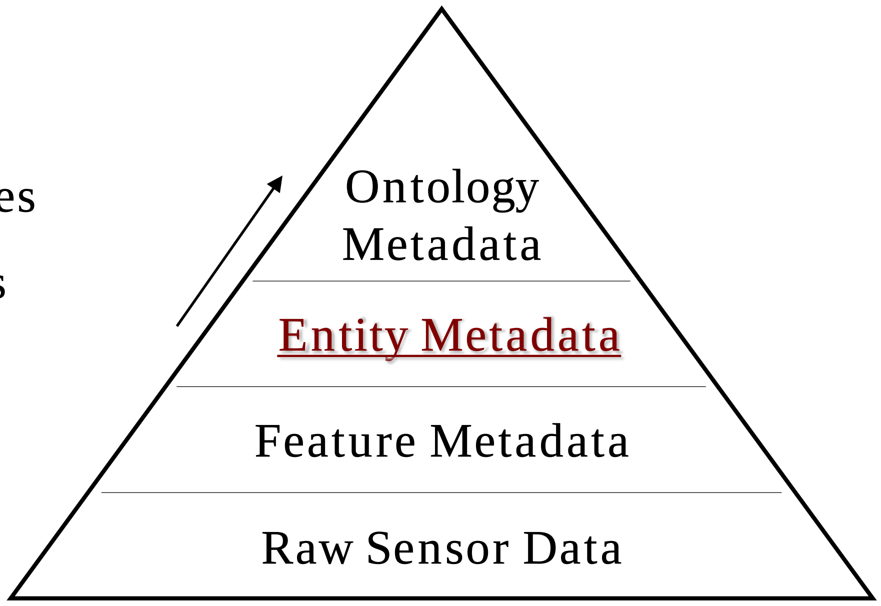
Solution Goal

1. Use O&M to model observations and measurements
2. Annotate sensor data with observation and measurement metadata
3. Store sensor metadata in XML database, and query with XQuery

Sensor Data Pyramid

Challenges

- Detect objects-events from features
- Annotate data with objects-events
- Store and query objects-events



Solution Goal

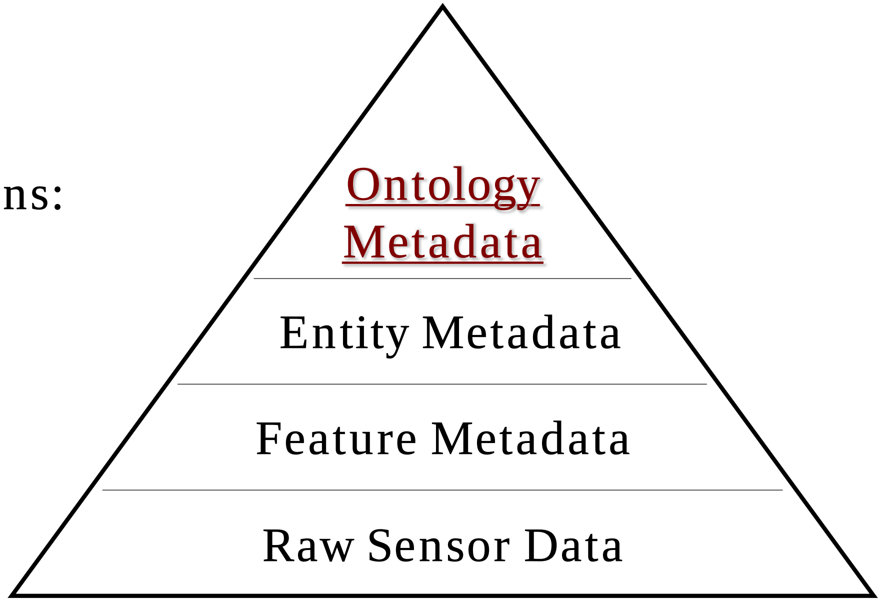
1. Build (or use existing) entity domain ontologies for objects and events
2. Extend SensorML with model-references to object-event ontologies
3. Annotate sensor data with object-event metadata
4. Store sensor metadata in XML database, and query with XQuery
5. Store object-event ontologies as RDF, and query with SPARQL

Sensor Data Pyramid

Challenges

Discover and reason over associations:

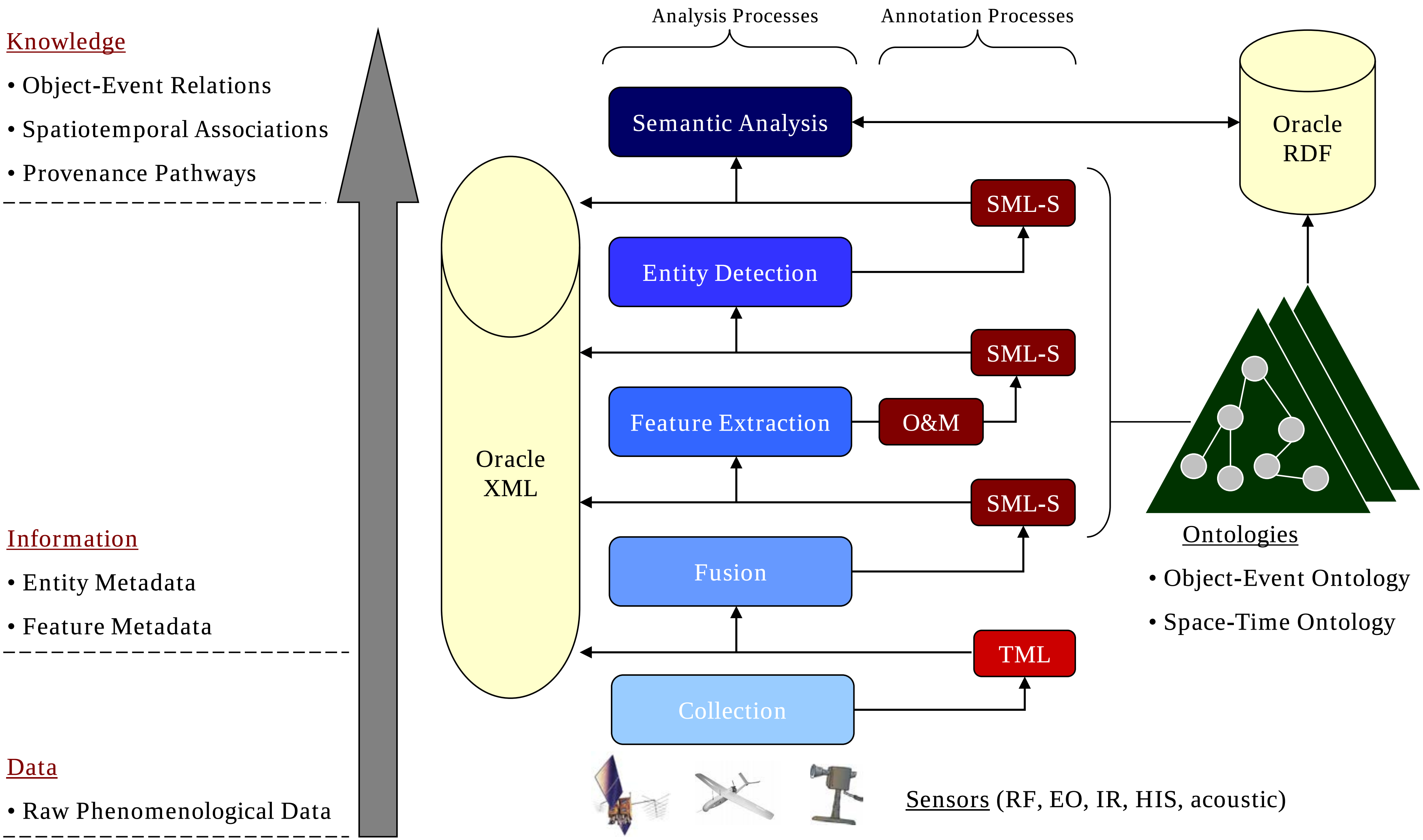
- objects and events
- space and time
- data provenance



Solution Goal

1. Query knowledge base with SPARQL
2. Object-event analysis to discover "interesting" events
3. Spatiotemporal analysis to track objects through space-time
4. Provenance Pathway analysis to track information through data life-span

Sensor Data Architecture



Presentation Outline

1. Motivating Scenario
2. Sensor Web Enablement
3. Sensor data evolution hierarchy
4. Semantic Analysis

Spatial, Temporal, Thematic Analytics

Three Dimensions of Information

Thematic Dimension: What

Temporal Dimension: When

North Korea detonates nuclear device on October 9, 2006 near Kilchu, North Korea

Spatial Dimension: Where

Motivation

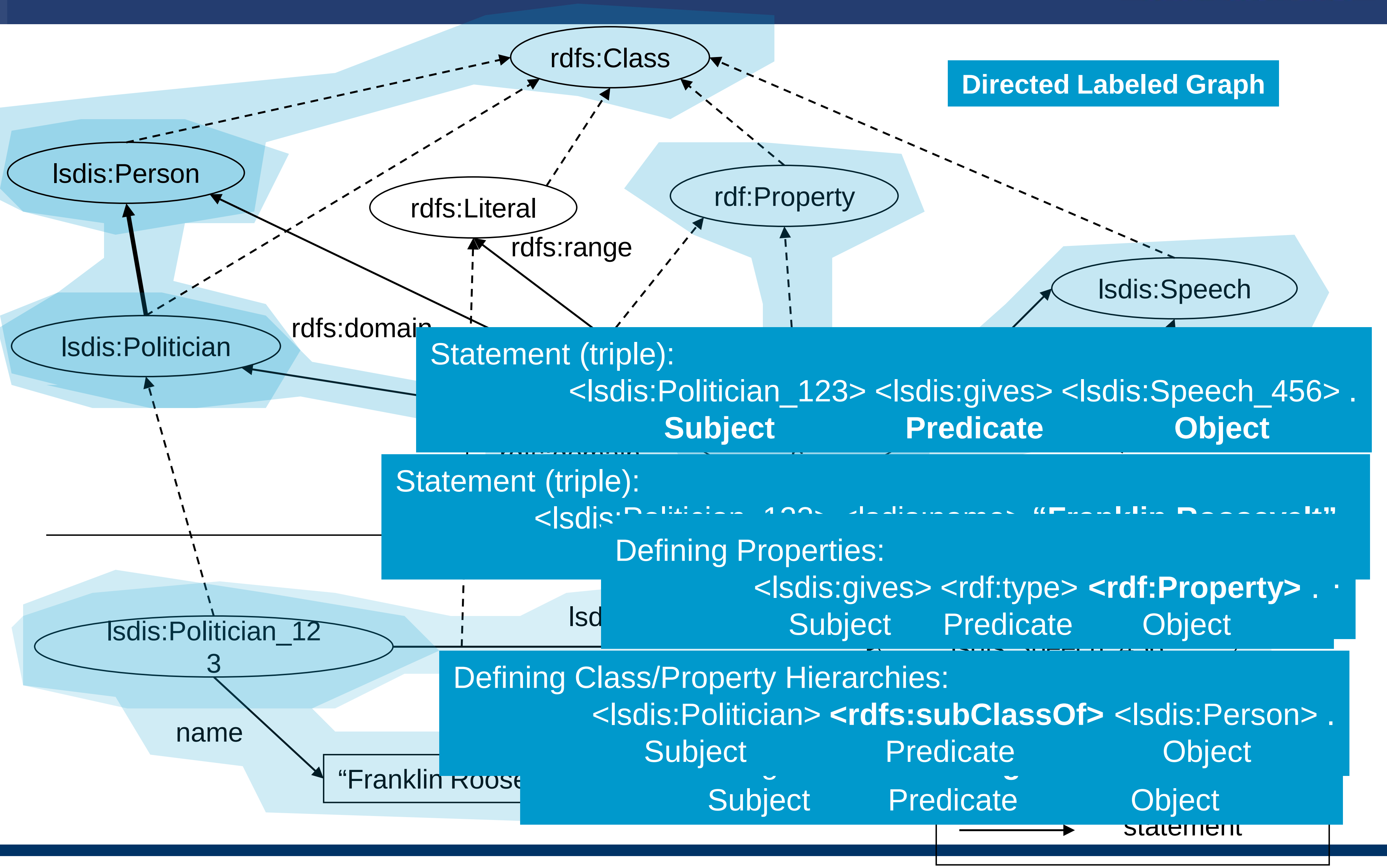
- **Semantic Analytics**
  - Searching, analyzing and visualizing semantically meaningful connections between named entities
- **"Connecting the Dots" Applications**
  - National Security, Drug Discovery, Medical Informatics
  - Significant progress with thematic data: query operators (semantic associations, subgraph discovery), query languages (SPARQ2L, SPARQLeR), data stores (Brahms)
- **Spatial and Temporal data is critical in many analytical domains**
  - Need to support spatial and temporal data and relationships

Value to Sensor Networks

- **Simple (Analyze Infrastructure):**
  - What types of sensors are available?
  - What sensors can observe a particular phenomenon at a given geolocation?
  - Get all observations for a particular geolocation during a given time interval.

- **Complex (More background thematic information):**
  - What do I know about vehicle with license plate XYZ123?
  - What do I know about the buildings (georeferenced) in this image?
  - Which sensors cover an area which intersects with a planned Military Convoy?

Directed Labeled Graph



Statement (triple):

<lsdis:Politician_123> <lsdis:gives> <lsdis:Speech_456> .

Subject **Predicate** **Object**

Statement (triple):

Defining Properties:

<lsdis:gives> <rdf:type> **<rdf:Property>** .

Subject Predicate Object

Defining Class/Property Hierarchies:

<lsdis:Politician> **<rdfs:subClassOf>** <lsdis:Person> .

Subject Predicate Object

Subject Predicate Object

Challenges

- **Data Modeling and Querying:**
  - Thematic relationships can be directly stated but many spatial and temporal relationships (e.g. distance) are implicit and require additional computation
  - Temporal properties of paths aren't known until query execution time … hard to index

- **RDFS Inferencing:**
  - If statements have an associated valid time this must be taken into account when performing inferencing
  - (x, rdfs:subClassOf, y) : [1, 4] AND (y, rdfs:subClassOf, z) : [3, 5] → (x, rdfs:subClassOf, z) : [3, 4]

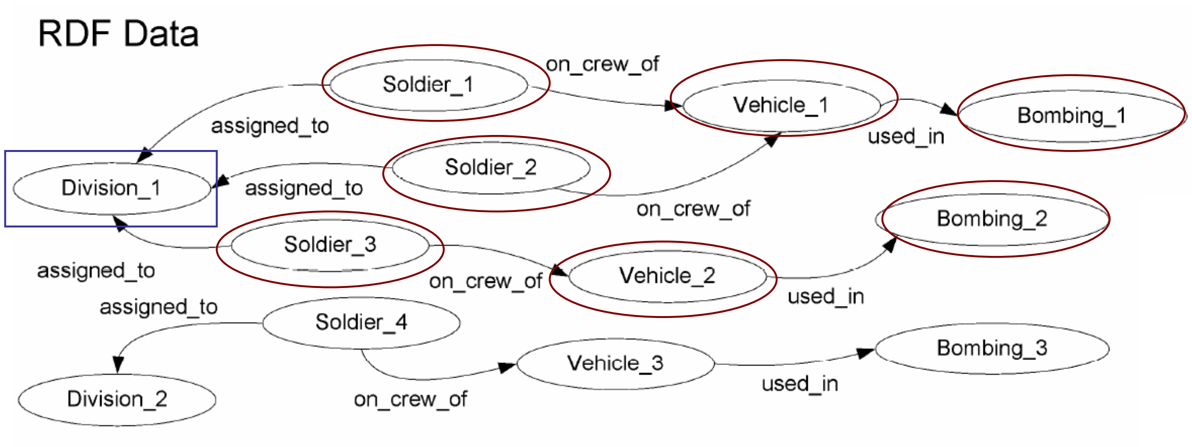
Work to Date

Kno.e.sis

- **Ontology-based model for spatiotemporal data using temporal RDF** [1]
  - Illustrated benefits in flexibility, extensibility and expressiveness as compared with existing spatiotemporal models used in GIS
- **Definition, implementation and evaluation of corresponding query operators using an extensible DBMS (Oracle)** [2]
  - Created SQL Table Functions which allow SPARQL graph patterns in combination with Spatial and Temporal predicates over Temporal RDF graphs

1. Matthew Perry, Farshad Hakimpour, Amit Sheth. "Analyzing Theme, Space and Time: An Ontology-based Approach", Fourteenth International Symposium on Advances in Geographic Information Systems (ACM-GIS '06), Arlington, VA, November 10 - 11, 2006
2. Matthew Perry, Amit Sheth, Farshad Hakimpour, Prateek Jain. "What, Where and When: Supporting Semantic, Spatial and Temporal Queries in a DBMS", Kno.e.sis Center Technical Report. KNOESIS-TR-2007-01, April 22, 2007

Example Graph Pattern



Graph Pattern

```
(?x assigned_to Division_1)
(?x on_crew_of  ?y)
(?y used_in     ?z)
```

Sample STT Query

```
select a from table (spatial_eval (‘(?a has_symptom ?b)
  (Chemical_X induces ?b)(?a fought_in ?c)’, ?c,
  ‘(?d member_of Enemy_Group_Y)(?d spotted_at ?e)’, ?e,
  ‘geo_distance(distance=2 units=mile)’));
```

Scenario (Biochemical Threat Detection): Analysts must examine soldiers’ symptoms to detect possible biochemical attack

Query specifies

Results

Operator (Exp. #)	Graph Pattern Type # Vars	# Triples	Queries	Avg. Result Size	Avg. Execution Time for each ontology (ms) Small	Medium	Large
T-Ext (1)	4	3	4	N/A	394	390	385
(2)	3	3	5	221	22	32	48
S-Ext (3)	4	3	3	N/A	360	350	365
(4)	3	3	3	100	22	30	67
T-Filter(5)	4	3	4	312	157	345	714
S-Filter (6)	4	3	3	331	173	192	374
T-Eval(7)	2/2	2/2	3	129	414	411	437
	2/3	3/3	3	220	306	195	268
S-Eval (8)	2/2	2/1	3	244	343	467	485
	2/2	2/3	3	209	251	385	457

Small: 100,000 triples Medium: 1.6 Million triples Large: 15 Million triples

Thank You.