Using Smartphones for Prototyping Semantic Sensor Analysis Systems

Hassan Issa, Ludger van Elst, Andreas Dengel

SBD 2016 Workshop – San Francisco – 01.07.2016

Technische Universität Kaiserslautern

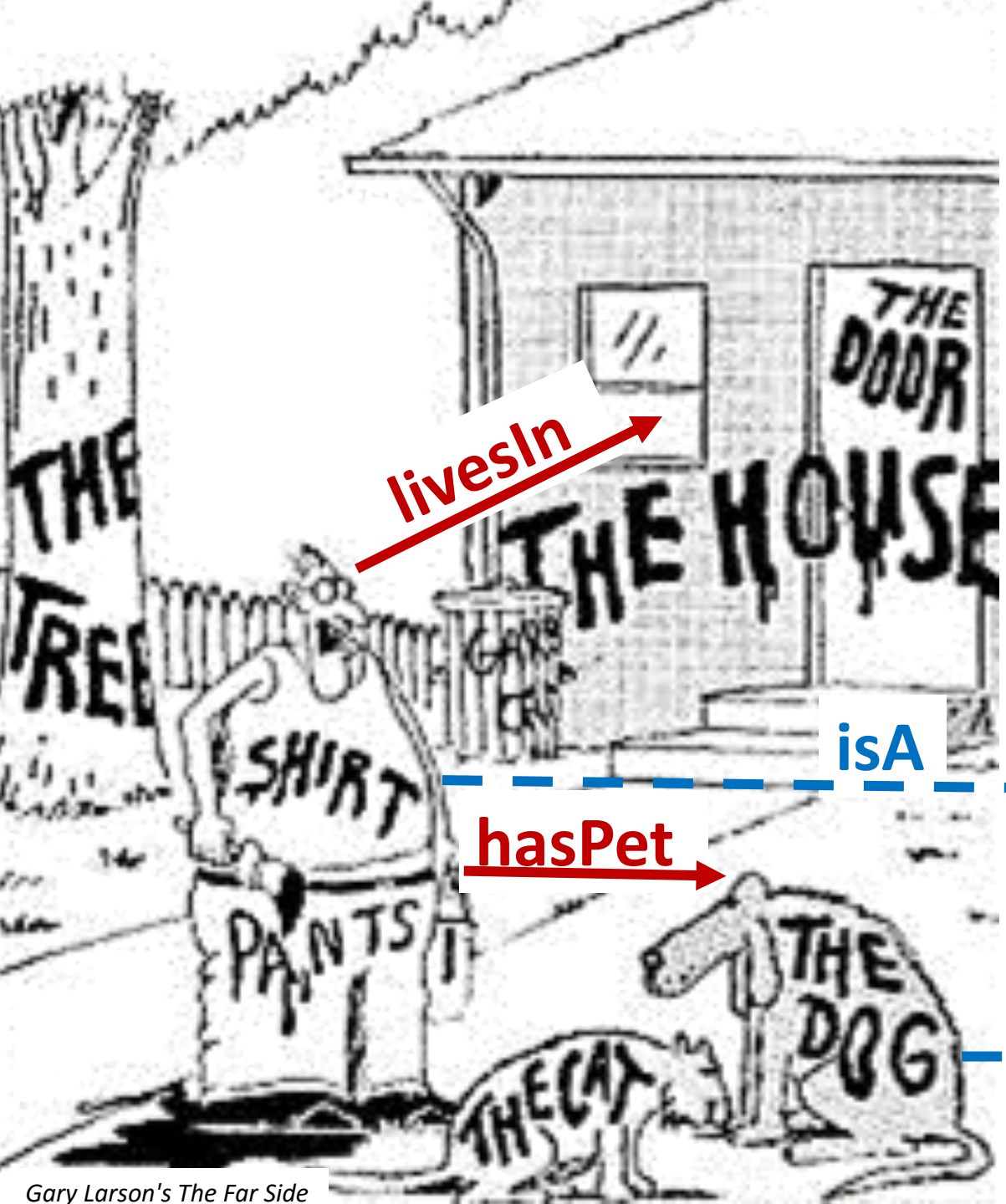
Over **5,000** sensors
in each engine

20 terabytes of data generated
per engine every **hour**

FRA to SFO

What if **all** these **sensors** go **online**?

What if **everything** goes **online**?



Gary Larson's The Far Side

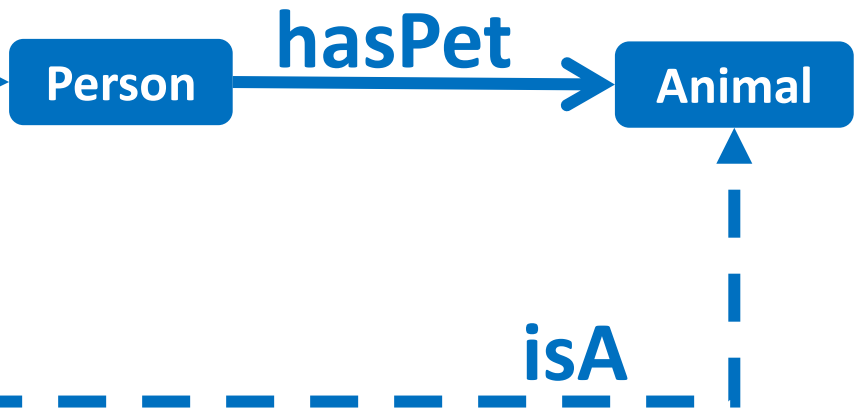
Semantic Web

Concrete Facts

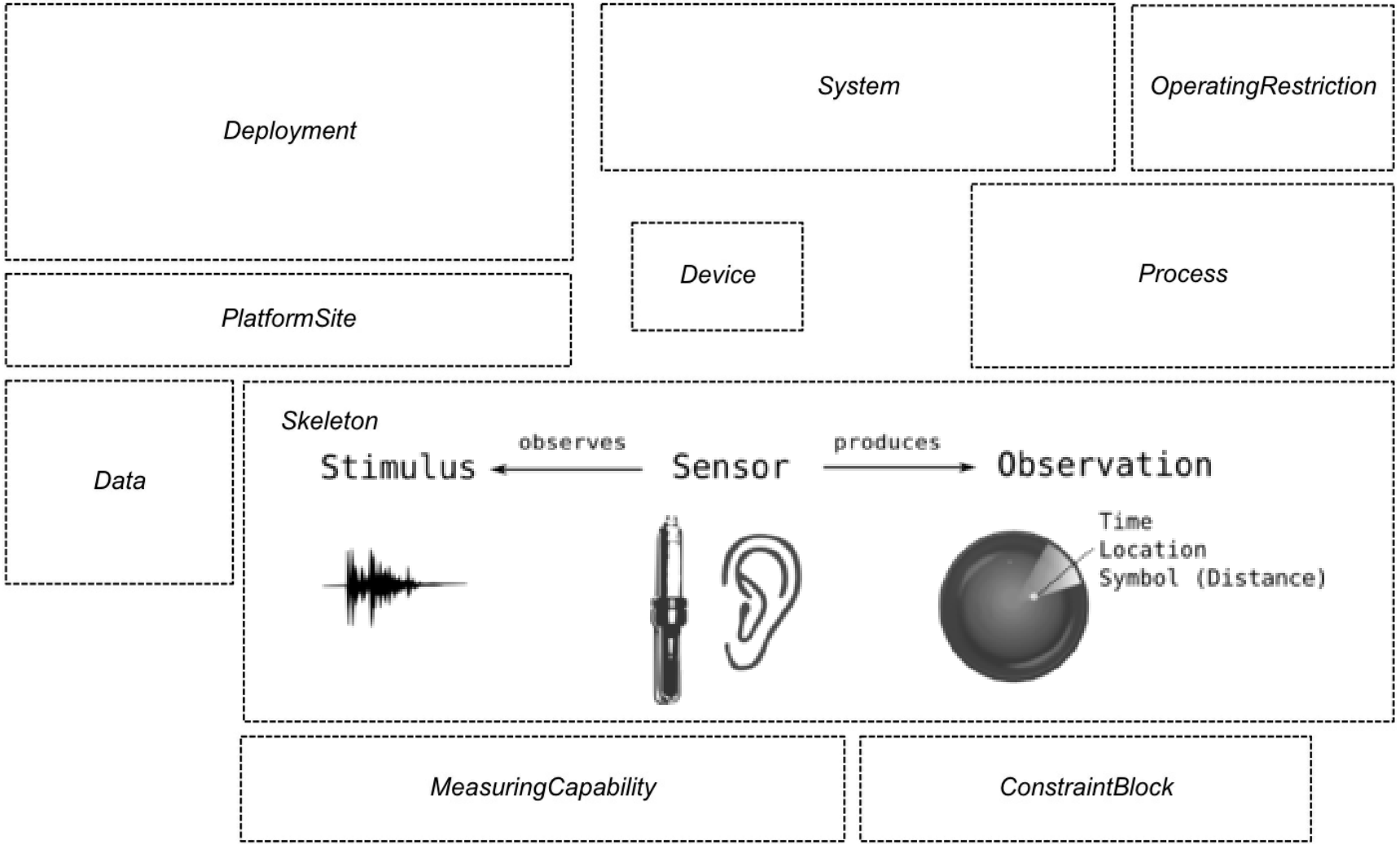
Resource Description Framework

General Knowledge

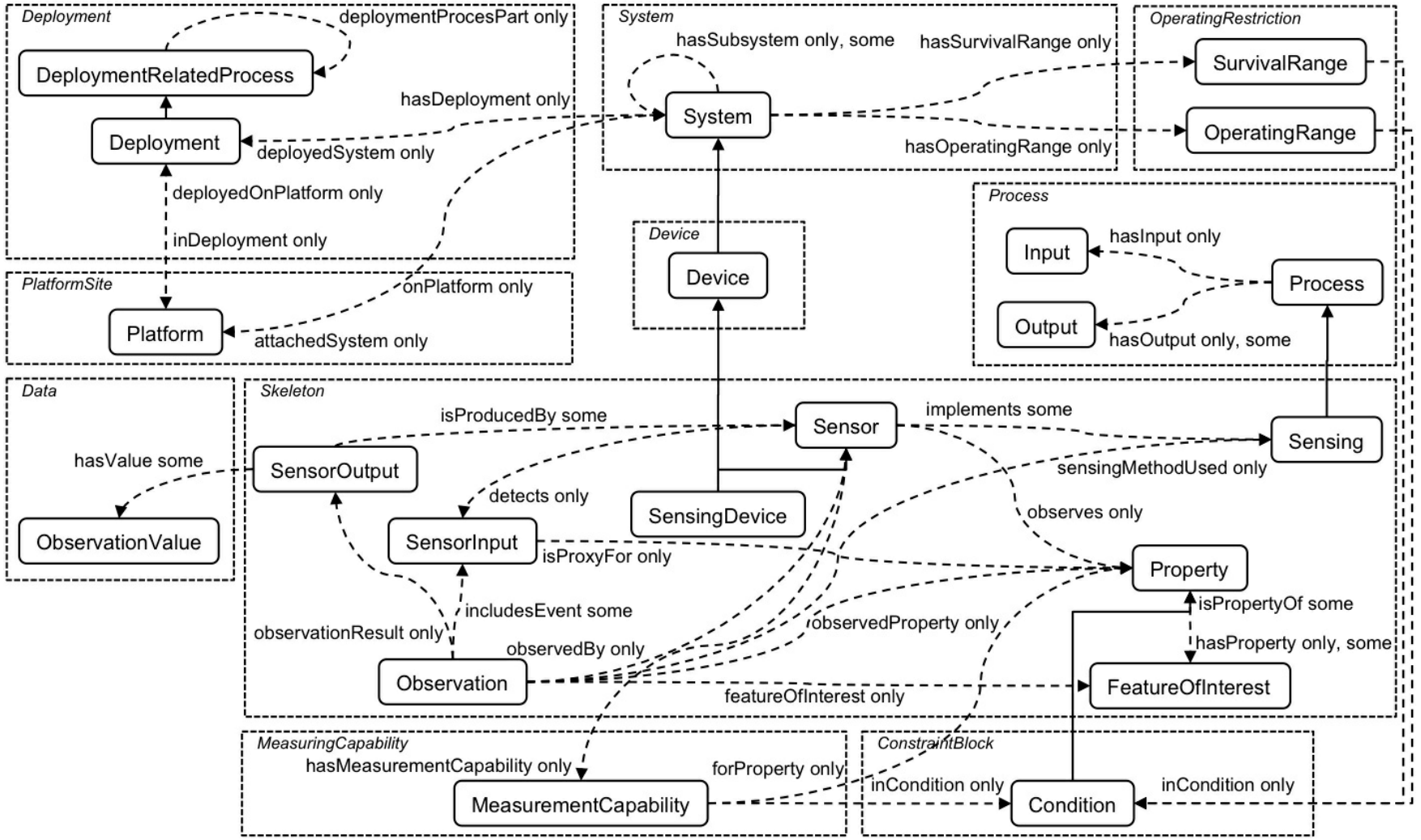
Web Ontology Language



Semantic Sensor Network Ontology



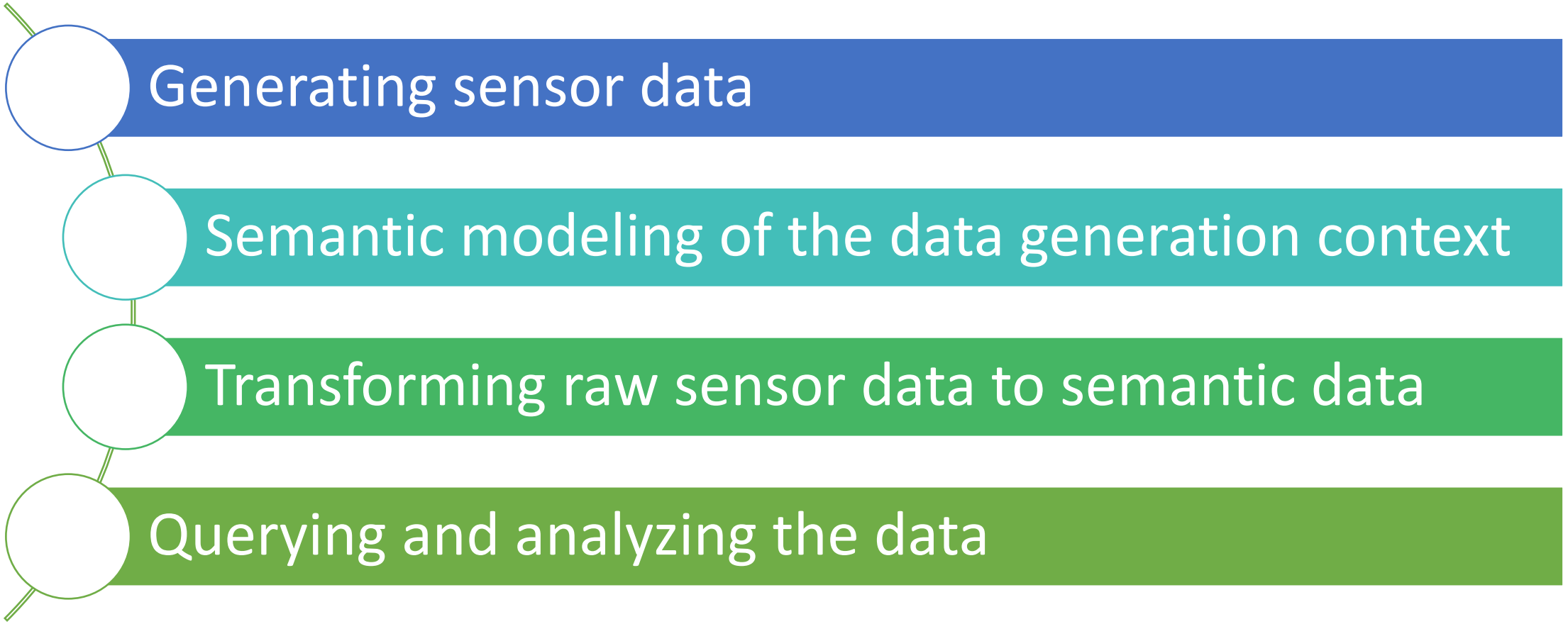
Semantic Sensor Network Ontology



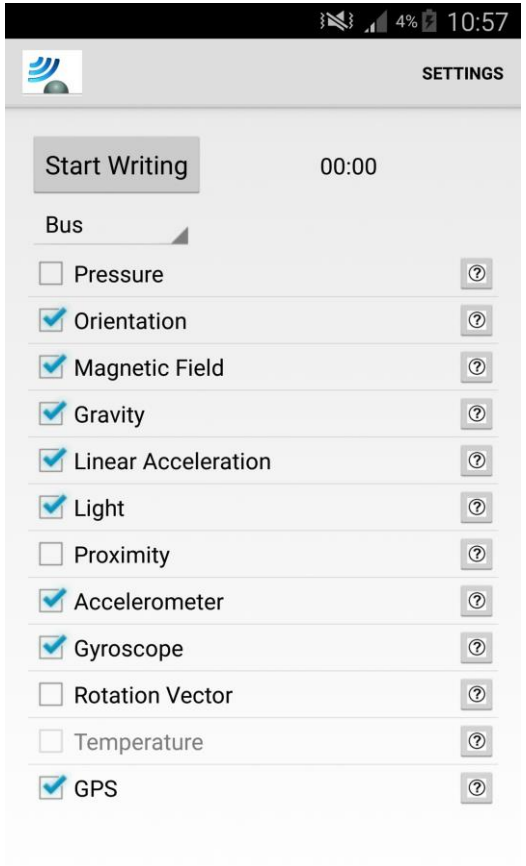
Advantages of Semantic Sensor Data

- Easy data integration
- Helps achieving autonomous processing and reasoning about sensor data
- Preserve data generation context
- Provides levels of abstraction

Prototype: Semantic Sensor Analysis System

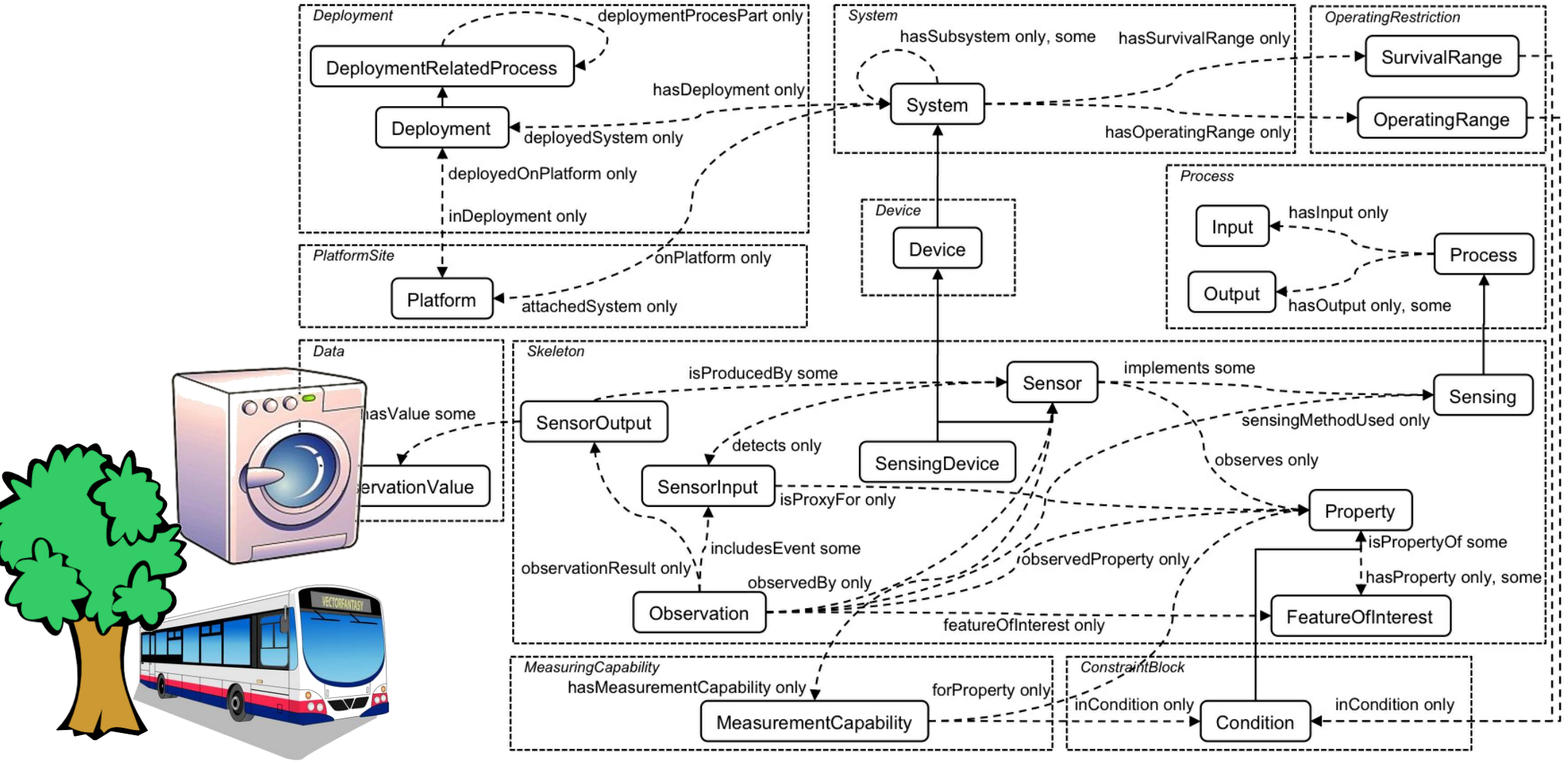


SensorTracker App

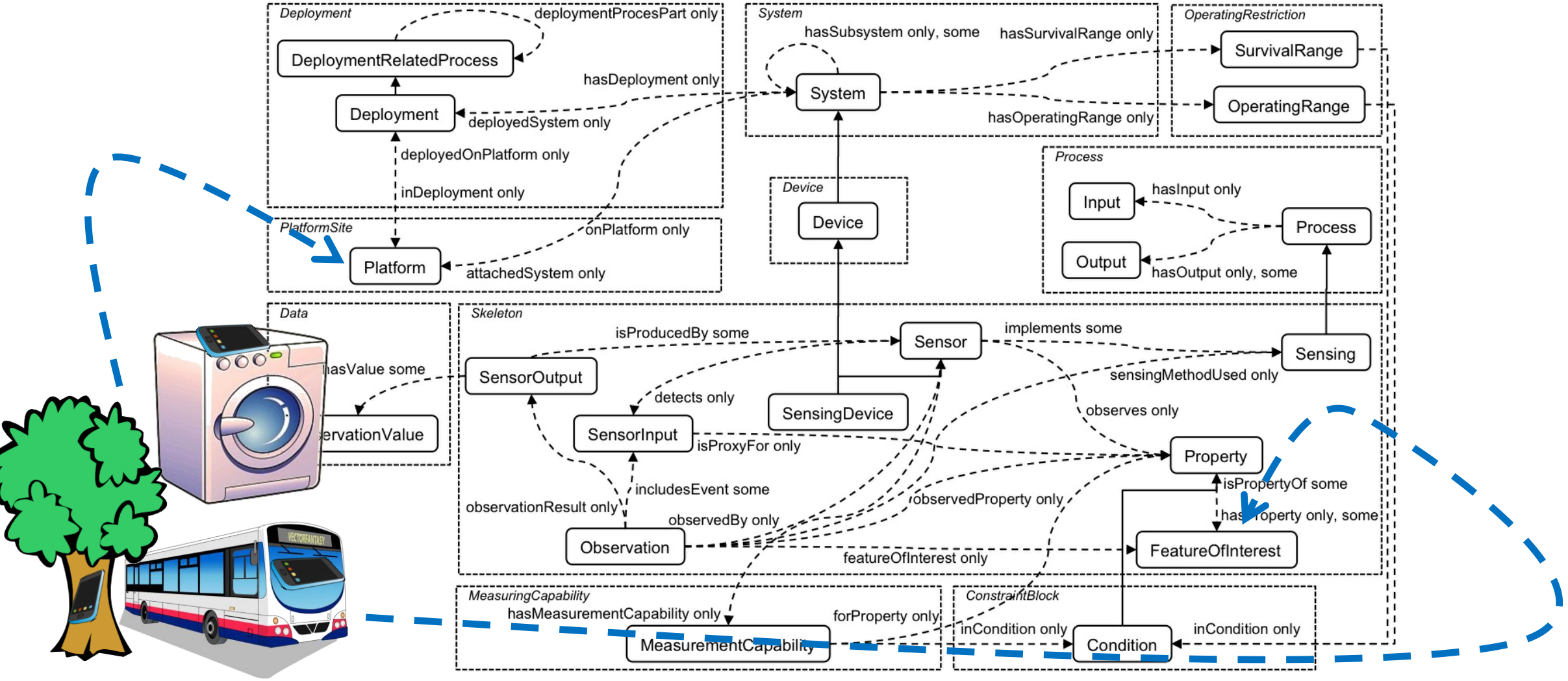


- Utilize smartphone sensors
- Easy deployment
- Data stored locally and/or transmitted over the internet
- Fused sensor data

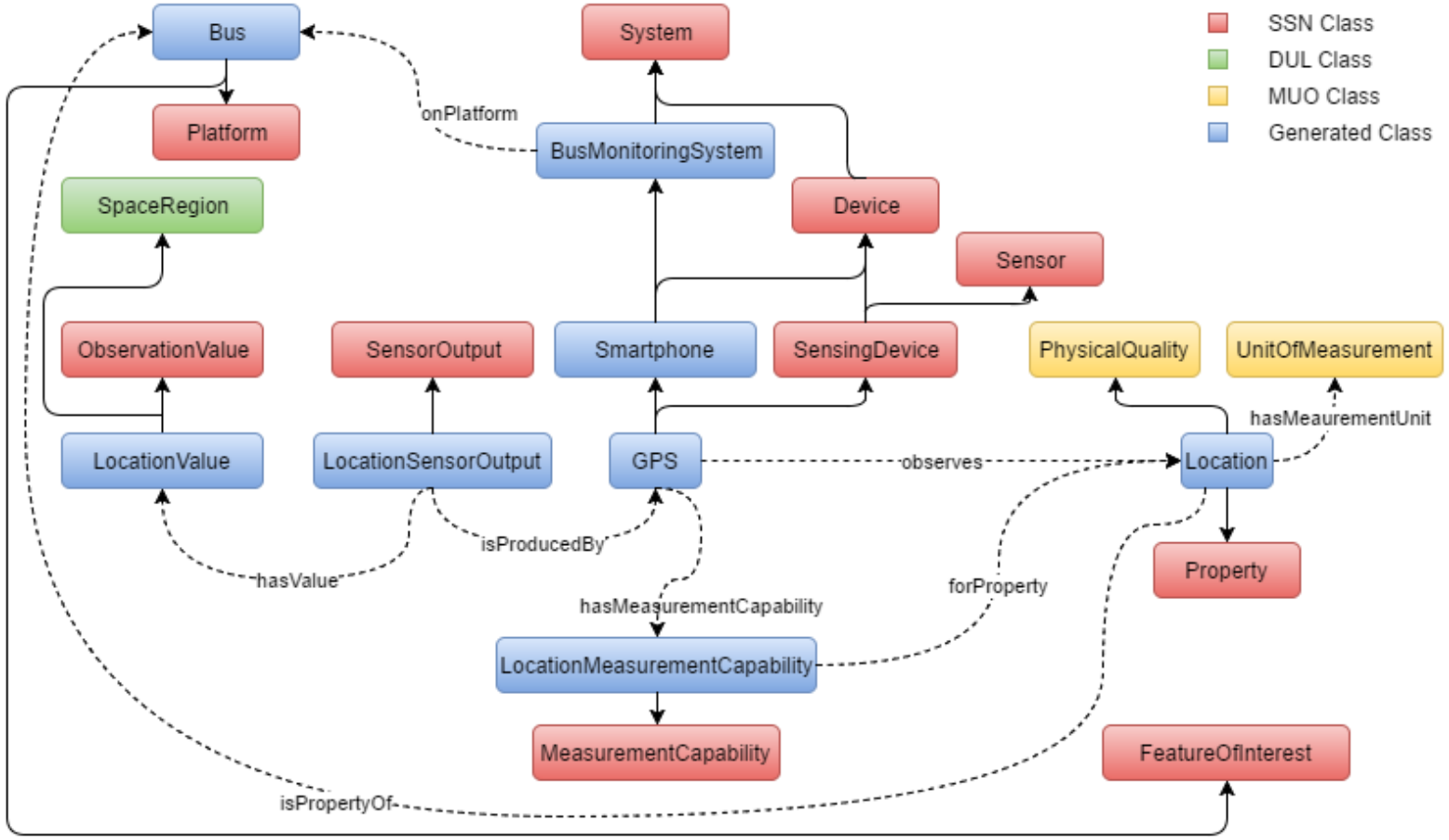
Generating an SSN-Based Ontology



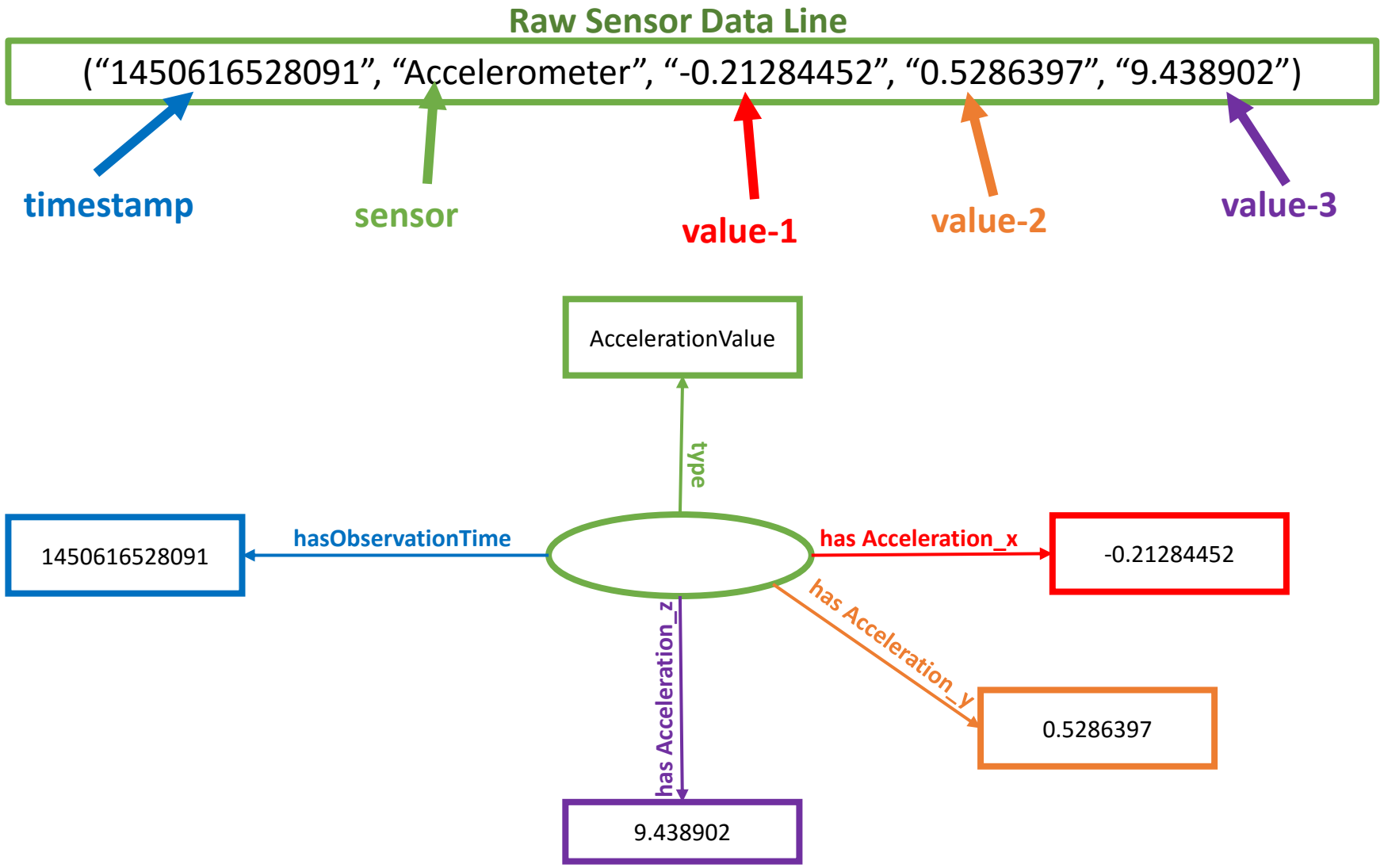
Generating an SSN-Based Ontology



Generating an SSN-Based Ontology



Raw Sensor Data to Semantic Data



Big Data Implementation

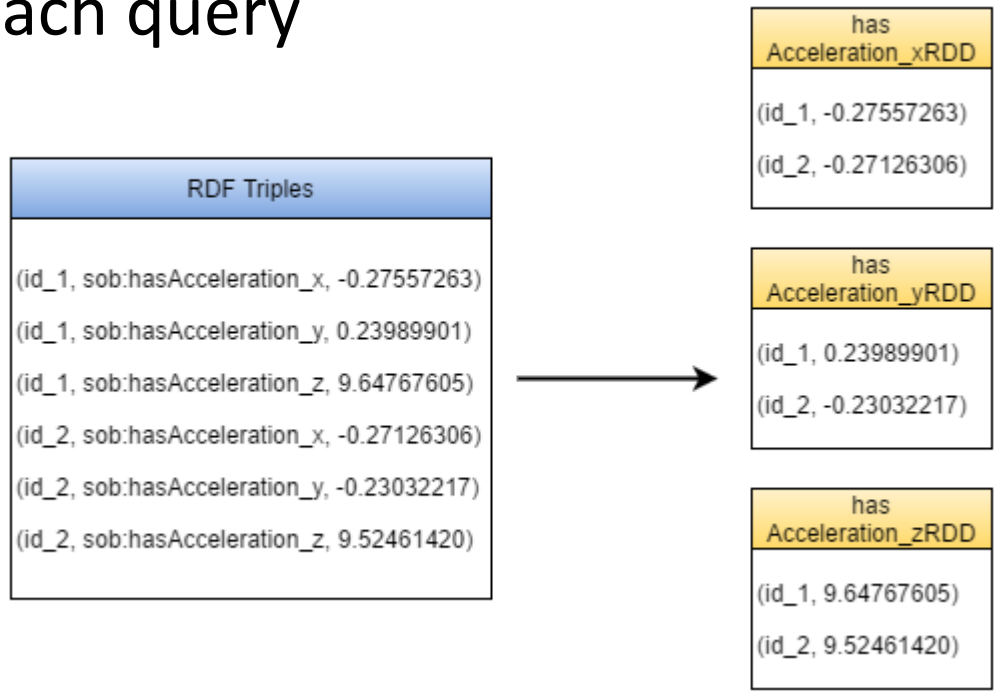
- Scale up to handle huge amounts of sensor data
- Using Apache Spark
- Distinguish TBox/ABox data
  - TBox data broadcasted to all nodes
  - Abox data distributed over cluster

TBox Encoding

- Base ontology is small
- Created locally on a single machine
- Two Tables created
  - Ontology Classes
    - Numerical id of a class
    - set of its sub-classes' ids
  - Property Classes:
    - Numerical id of property
    - IDs of domain and range classes
    - Ids of sub-properties

ABox Encoding

- Raw sensor data transformed into triples
- Triples are stored in separate RDDs for each property
- Not all triples are loaded on each query



Querying Semantic Sensor Data

- Use Spark operations for analysis
- SPARQL queries are transformed into a set of spark operations
  - Map
  - Filter
  - Join
  - …

```
SELECT ?lat ?lon
    WHERE{
?a rdf:type sob:LocationValue.
?a sob:hasLatitude ?lat.
?a sob:hasLongitude ?lon.
}
```

SPARQL

```
(TypeRDD
 .filter(lambda (nodeID,typeID):
        equalsType(typeID,"sob:LocationValue") )
 .join(hasLatitudeRDD)
 .join(hasLongitudeRDD)
 .map(lambda (nodeID,(typeID,lat,lon)): (lat,lon))
).collect()
```

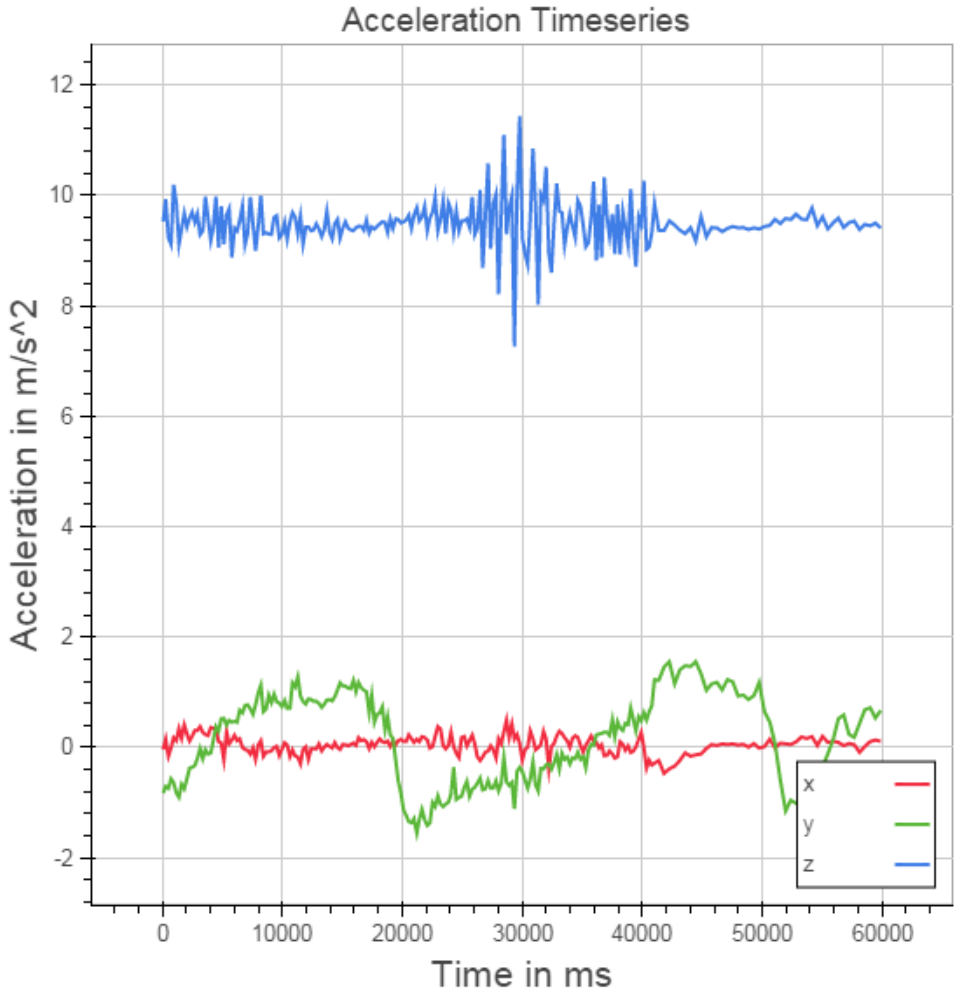
SPARRK

Street Quality Assessment Application

- Smartphone deployed in a public transport bus
  - 8 days
  - 1600km
  - 14+ million records

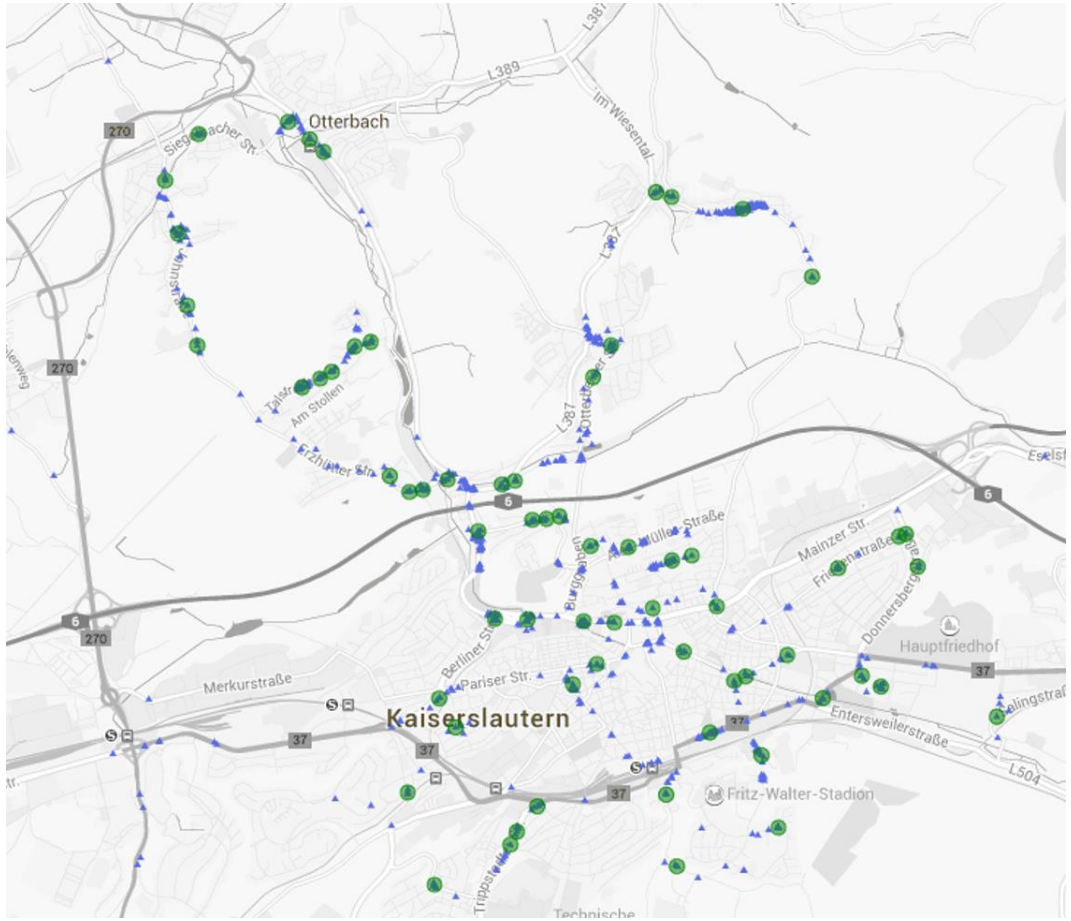
  (Germany: 153+ billion records/day)

Street Anomalies Detected



Spike of at least 1.8m/s^2 in a 2-second time frame

Street Anomalies' Locations



Clustering using DBSCAN

Conclusions

- Introduced a prototype for sensor analysis systems
  - Smartphone used for data collection
  - SSN-based ontology generated to describe the sensor setup
  - Raw sensor data transformed to semantic data
  - Spark used for data transformation and analysis
- System is scalable and can integrate data from different sources
- Street quality assessment use case

Future Work

- Collect ground truth and evaluate street anomalies results
- Create a complete ontology for bus networks according to German open data
- Evaluate and optimize storage scheme followed

Thank you.