INSTITUTE FOR COMMUNICATION SYSTEMS

5G INNOVATION CENTRE

UNIVERSITY OF SURREY

A Data-driven Approach for Internet of Things Applications: Methods and Case Studies

Suparna De

30th October, 2017

University of Granada, Spain

CITIC-UGR

SEFORI 2017

VII SEMINARIOS DE FORMACIÓN PARA LA INVESTIGACIÓN EN TIC

Outline

- Internet of Things: an Introduction
- Data Processing pipeline
  - Data Sources
  - Data Modelling: a Semantic Approach
  - Data Search and Retrieval
  - Data Analysis - Reasoning Methods
    - Case Studies
  - IoT Application Domains
- Open Research

Internet of Things (IoT): an Introduction

THE INTERNET OF THINGS

- Term coined by Kevin Ashton in 1999.
- Interconnection of objects to computers with self-configuring capabilities
- Main enablers:
  - sensors and actuators embedded in physical objects
  - RFID and sensor technology enable computers to observe, identify and understand the world
- Drivers:
  - things-to-things communications
  - integration of things data with applications

Low-cost Sensors are becoming prevalent

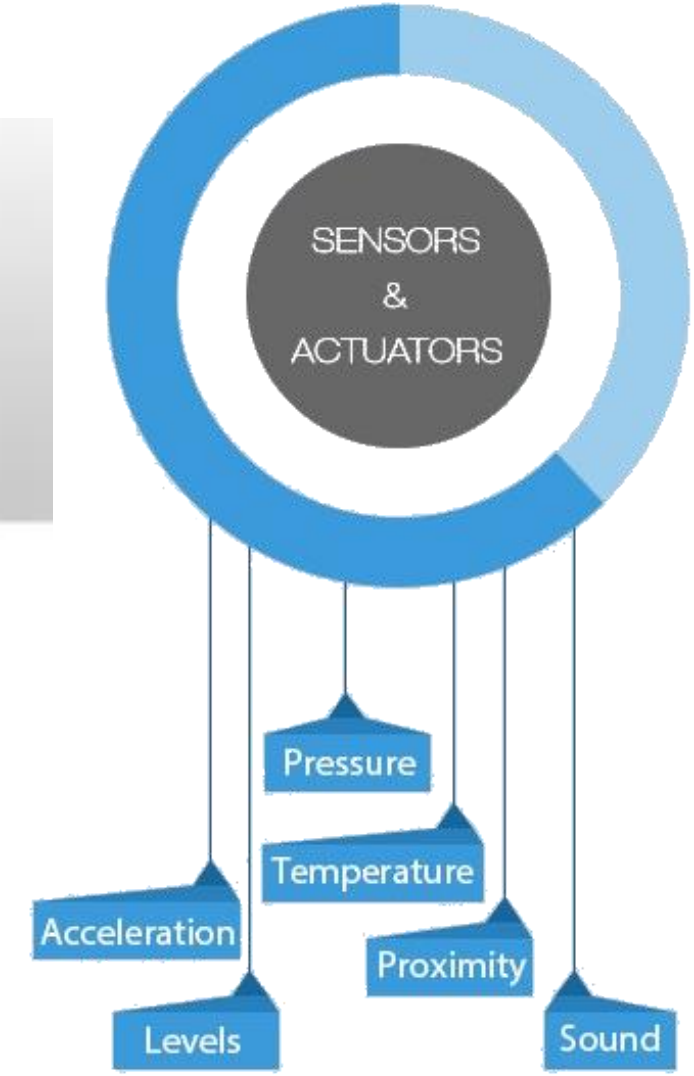
Environment sensors

Utility consumption sensors

Dynamic Tags

UPCODE Ltd.

TAGIT SMART



More parts of life are getting connected…

Cities
Public transport
Consumer goods
Smart Homes

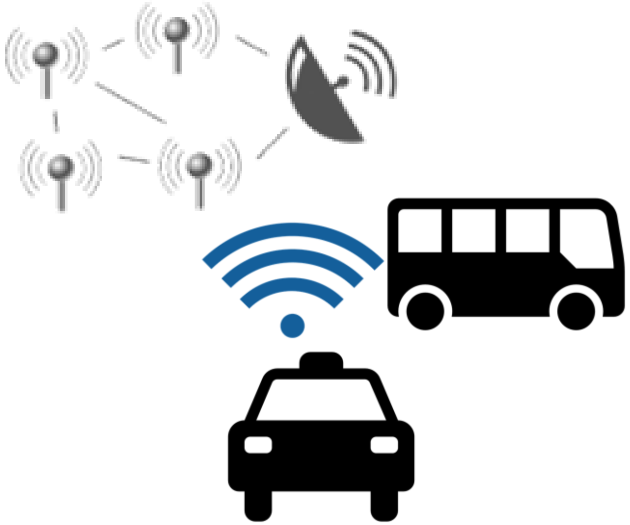


Image courtesy: 1. Exigent Networks; www.exigentnetworks.ie
2. SmartCitiesCouncil

IoT Drivers

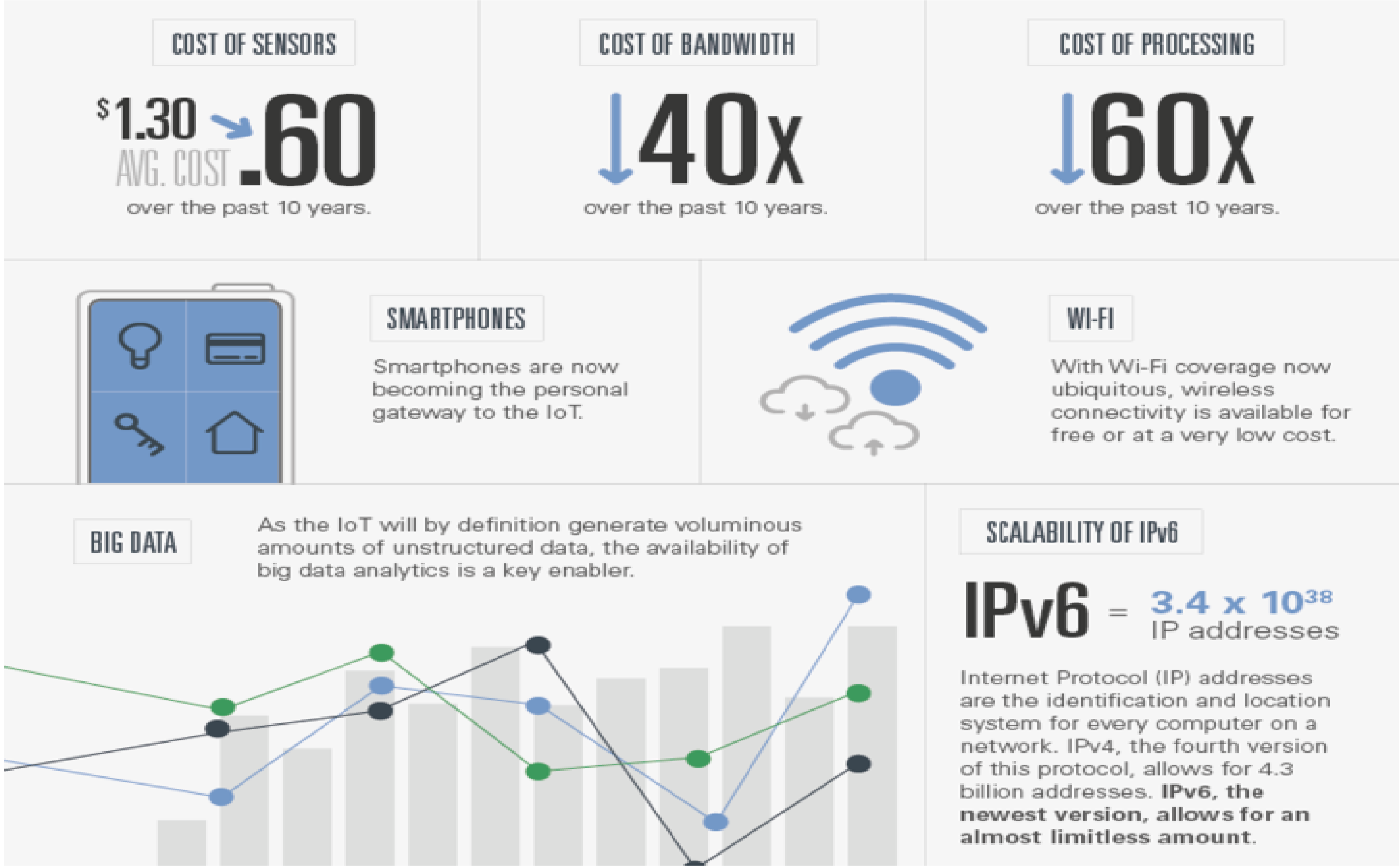
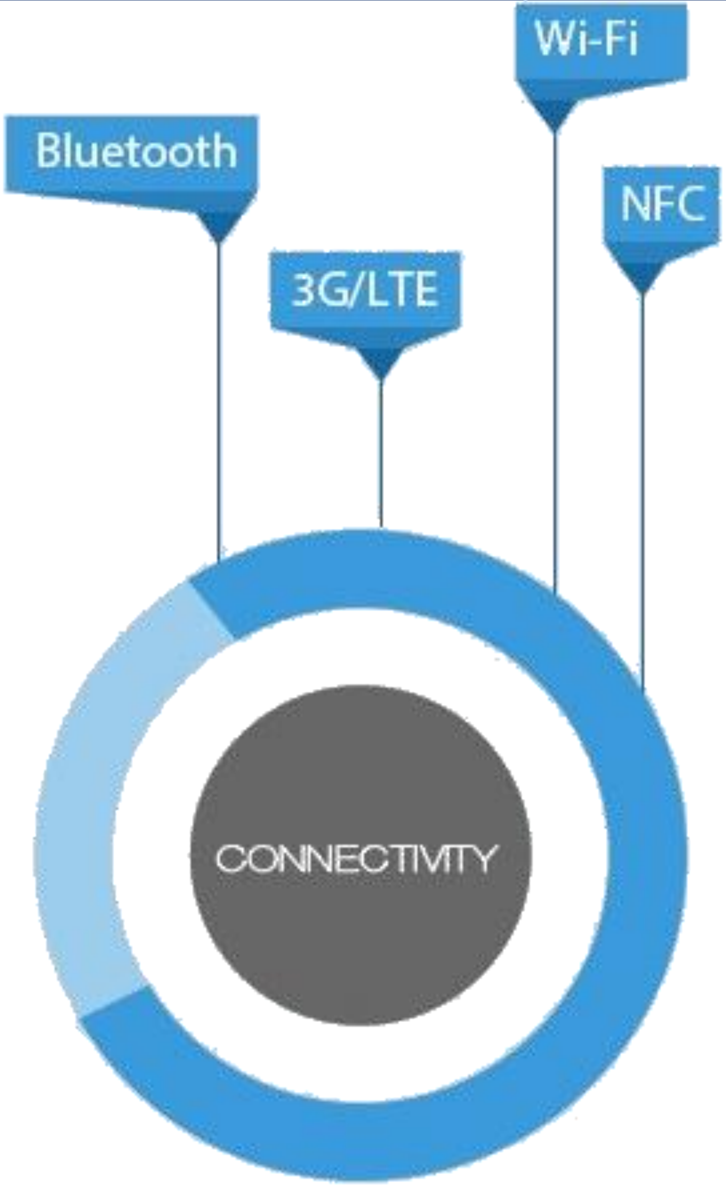


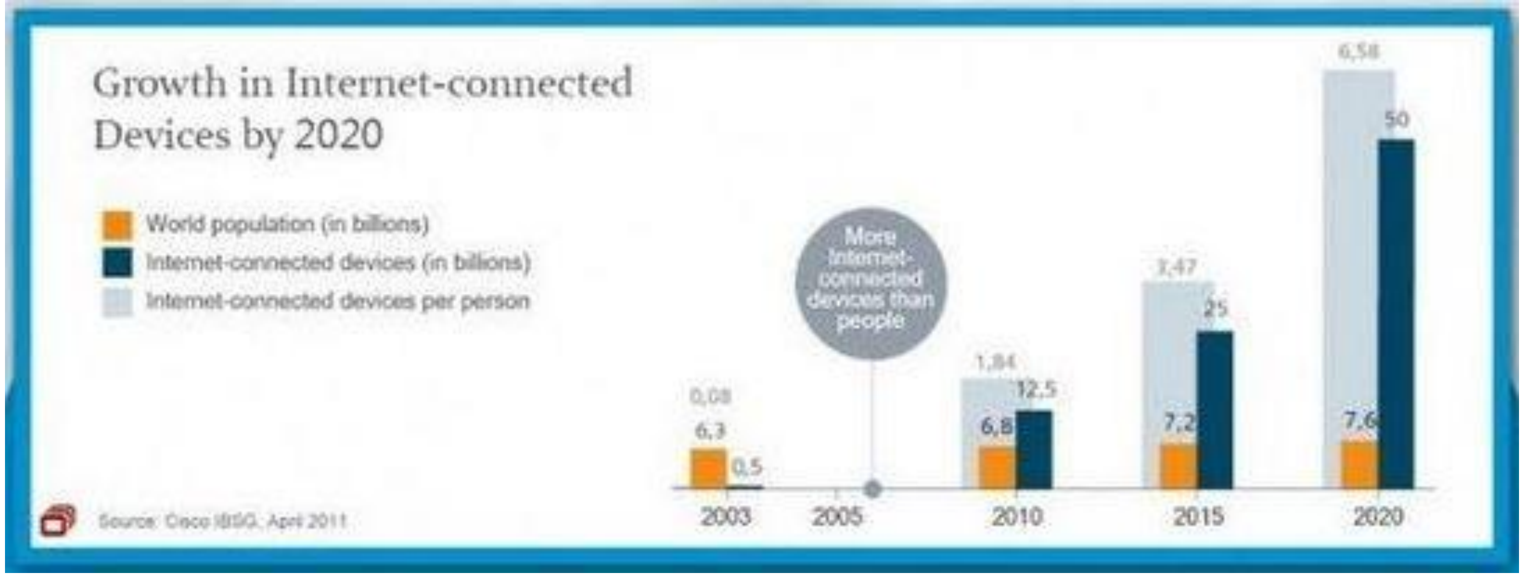
Image ©: Goldman Sachs Global Investment Research

From IoT to the Web of Things (WoT)

- Connecting "Things" to the Web for:
  - access
  - description and discovery
  - resource directories
  - security
- Typical connectivity solutions:
  - Constrained Application Protocol (CoAP)
  - Lightweight HTTP



IoT in numbers…



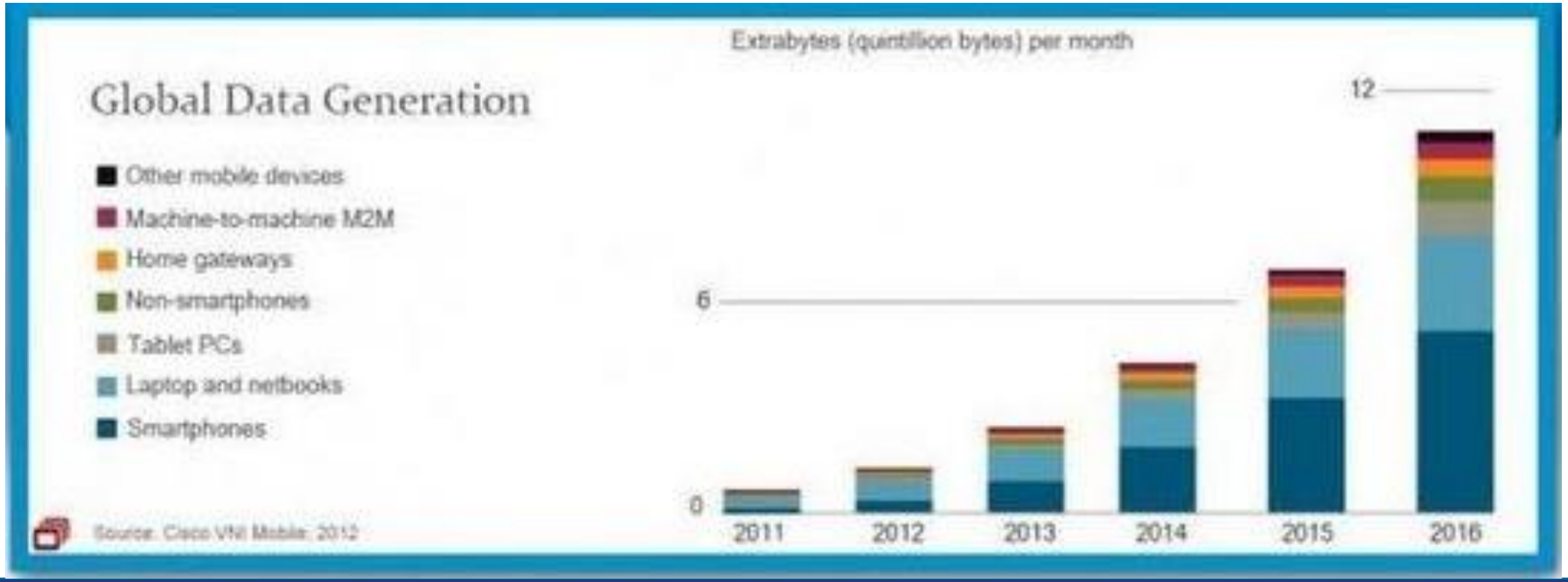


Image courtesy: Exigent Networks; www.exigentnetworks.ie

IoT: the case for a Data Perspective

- Abstractions of high-dimensional, high-volume data generated by heterogeneous devices
- Associate data with context information
- Data fusion through application of data analytics and reasoning techniques
- Example Applications:
  - Analyse road, environment (pollution) conditions with real-time location information (proximity) to recommend events and venues
  - Adjust traffic signal timings based on vehicle and cyclist arrival data
  - Efficient waste management using FMCG (fast moving consumer packaged goods) lifecyle information from smart tags

Challenges in IoT realization

- Many devices do not speak the same language and cannot exchange data across different gateways and smart hubs
- Things' data may have a defined structure in a known format, e.g. JSON, CSV, XML
- But data models adopted are different and not compatible
- Different units and context representation

Interoperability Challenge

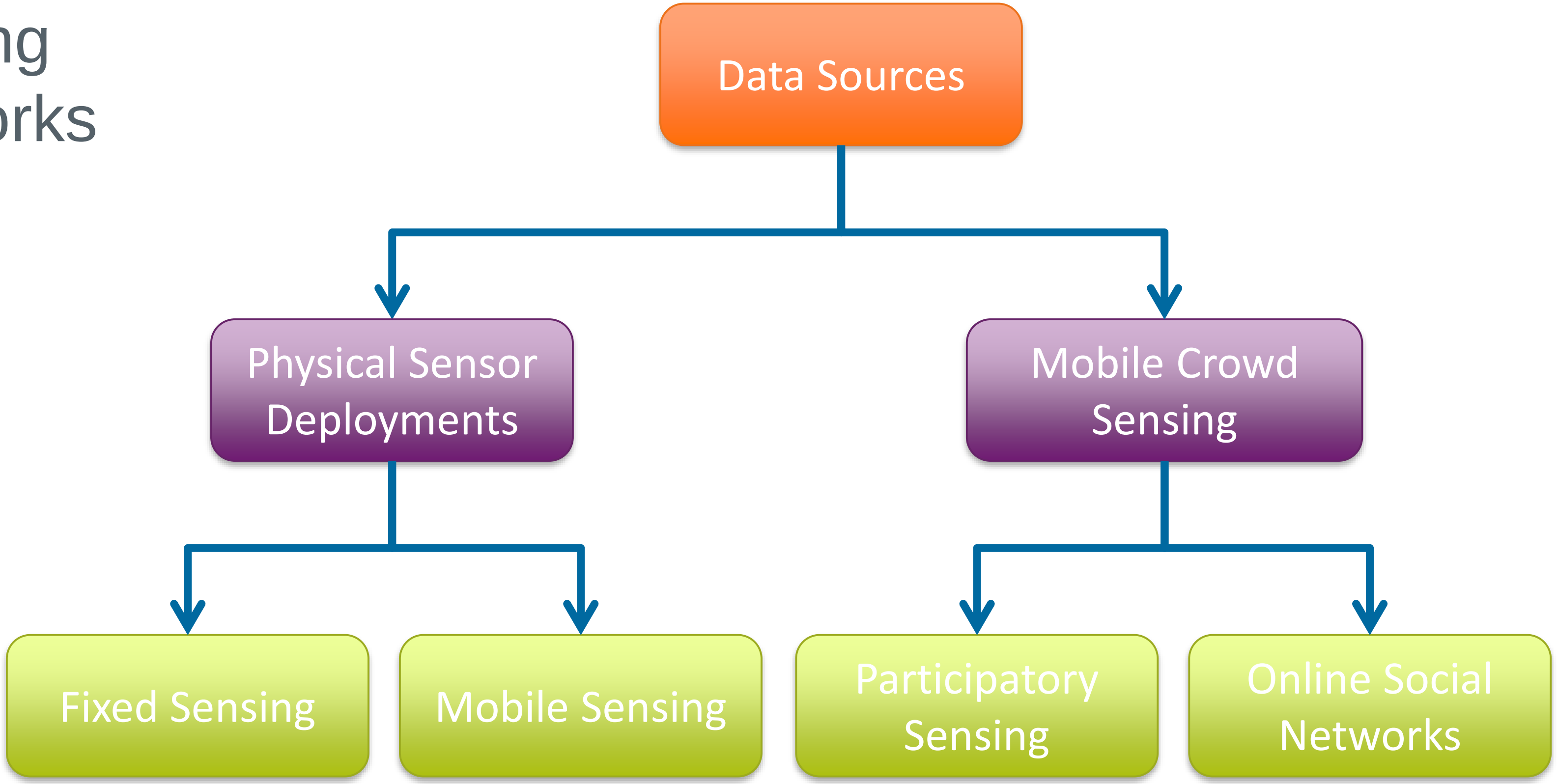
The Data Lifecycle

- Data Collection
  - Identification and connection to data sources
  - Physical and social world data sources
  - Data virtualisation
    - Data modelling
    - Schema adaptors
- Data Management
  - Data indexing
  - Metadata, data storage and retrieval
- Data Processing
  - Missing data estimation
  - Redundancy filtering, pre-sorting…
- Data Analysis
  - Reasoning mechanisms
  - Data fusion
- Applications

Data Sources

Data Source taxonomy

- Physical sensor deployments
  - Fixed Sensing
  - Mobile sensor nodes
- Mobile crowd sensing
  - Participatory sensing
  - Online social networks



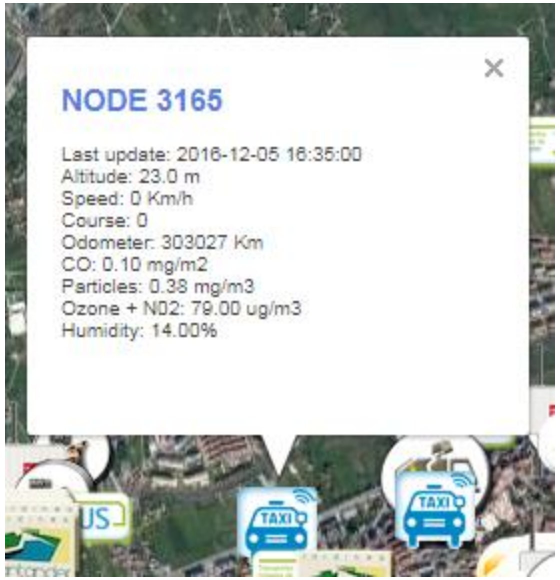
Physical Sensor Deployments

- Fixed Sensing
  - Fixed installations – static location configuration of deployed sensors
  - O&M data ~ continuous time series, resolution dependent upon sampling rate
  - Typical deployments in urban areas, smart homes, ITS solutions
  - Structured data, typically in JSON/CSV/XML formats
  - Examples:
  - London Air Quality Network (www.londonair.org.uk/London)
    - Air pollution sensors: CO, NO2, O3, PM10, PM2.5, SO2
    - Data sampled every 15 minutes
    - 4 location types: roadside, suburban, urban background, industrial
    - APIs for accessing data in XML or JSON; historical downloads in CSV
  - Smart Santander (http://maps.smartsantander.eu/)
    - Environmental monitoring: temperature, CO, noise, light
    - Traffic monitoring: traffic volume, road occupancy, vehicle speed, queue length
    - Agriculture monitoring: moisture, temperature, humidity…
  - Water Distribution Networks…

London Air

Physical Sensor Deployments (2)

- Mobile Sensing
  - O&M data ~ typically frequently updated, timestamped and structured (FUTS) data
  - Each sampling data point associated with a distinct location tag
  - Data typically accessible in a known structured format
  - Not obtained at successive, equally-spaced time points
  - Typical deployments for urban monitoring
  - Opportunity for large-scale environmental monitoring
  - Structured data, typically in JSON/CSV/XML formats
  - Examples:
    - Smart Santander
      - » http://maps.smartsantander.eu/
      - » Sensors attached to public vehicles
      - » Environmental monitoring:
      - » temperature, CO, noise, light
    - Madrid
      - » Pollen sensors on public buses




```
{
  "id":"3165",
  "latitude":"43.464900",
  "longitude":"-3.824330",
  "title":"bus3165",
  "image":"http:\/\/lira.tlmat.unican.es\/SmartSantander\/map\/images\/taxi.png",
  "content":"<div class='googft-info-window'\n    style='font-family:
  sans-serif; font-size: 10px;width: 200px; height: 18em ; overflow-y: auto;'
  ><table width='100%' border='0'>\n  <tr>\n    <td valign='top'><h2 style=
  'color: #5080e1'>NODE 3165<\/h2>Last update: 2016-12-05 16:35:00<br>Altitude:
  23.0 m<br>Speed: 0 Km\/h<br>Course: 0<br>Odometer: 303027 Km<br>CO: 0.10
  mg\/m2<br>Particles: 0.38 mg\/m3<br>Ozone + NO2: 79.00 ug\/m3<br>Humidity:
  14.00%<\/td>\n    <td valign='top'><\/td>\n  <\/tr>\n<\/table><\/div>",
  "tags":"BUS"
}
```


Image courtesy: Smart Santander testbed; http://maps.smartsantander.eu

Mobile crowd sensing

- Participatory Sensing
  - Smartphone accompanied citizens forming sensing networks for local knowledge gathering
  - Involves explicit participation
  - Made possible through dedicated apps or hardware carried by citizens
  - Examples:
    - Congested road and traffic incident detection
      - » Arduino boards in cars: speed and position of the car
      - » Environmental monitoring: temperature, CO, noise, light
    - Melbourne: noise heatmaps of city
      - » Complement noise data from fixed sensors, perceptions of noise and urban sounds

Mobile crowd sensing (2)

- Online social networks
  - Immediacy of social network messages: rich source of city-information
  - Data amenable to mining
  - Can provide semantic context to physical sensing data
  - Examples:
    - Twitter
      - » 140 character tweets
      - » Wide adoption – 500 million users worldwide
      - » Both streaming and RESTful API
      - » User perception of pollution, representation term mining for traffic incidents
    - Foursquare
      - » Location-based social network
      - » Users check-in to a venue
      - » Data: time, type, user details, venue details (name, location, category)
      - » Modality and format allows direct manipulation through statistical methods and integration with (numeric) physical sensing data

FOURSQUARE

Data Modelling

Semantics for IoT Resources, data

- Semantics: machine-processible metadata (tagging)
- Semantic languages
  - Web Ontology Language (OWL), Resource Description Framework (RDF)
- Structured, common platform for Things and data representation [*Data Modelling*]
- Achieving: ontologies for IoT "Things"
  - Resource model
    - Gateway, sensors, processing resources
  - Entity model
    - Physical world objects
    - Features of interest for each entity
  - Service model
    - IoT services and interfaces
  - Observation and Measurment (O&M) data
- Machine interpretation of relationships and hierarchies

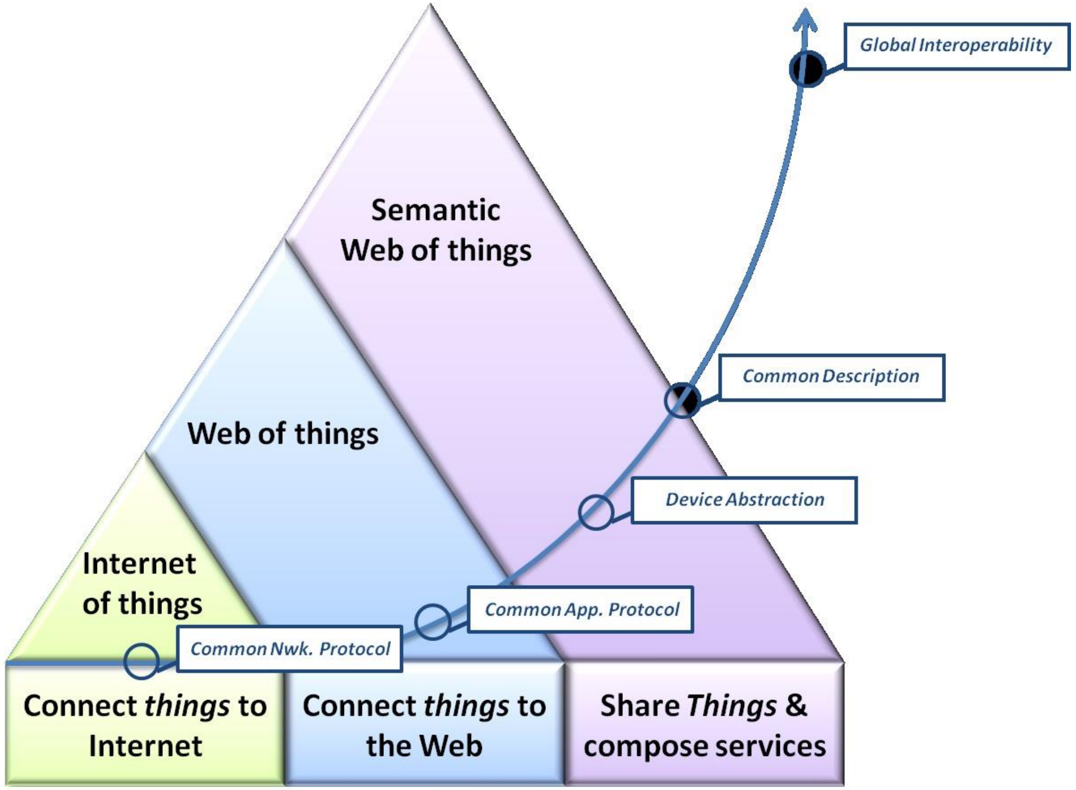


Image ref: Jara *et al*. Semantic web of things: An analysis of the application semantics for the iot moving towards the iot convergence. Int. J. Web Grid Serv., 10(2/3):244–272, April 2014

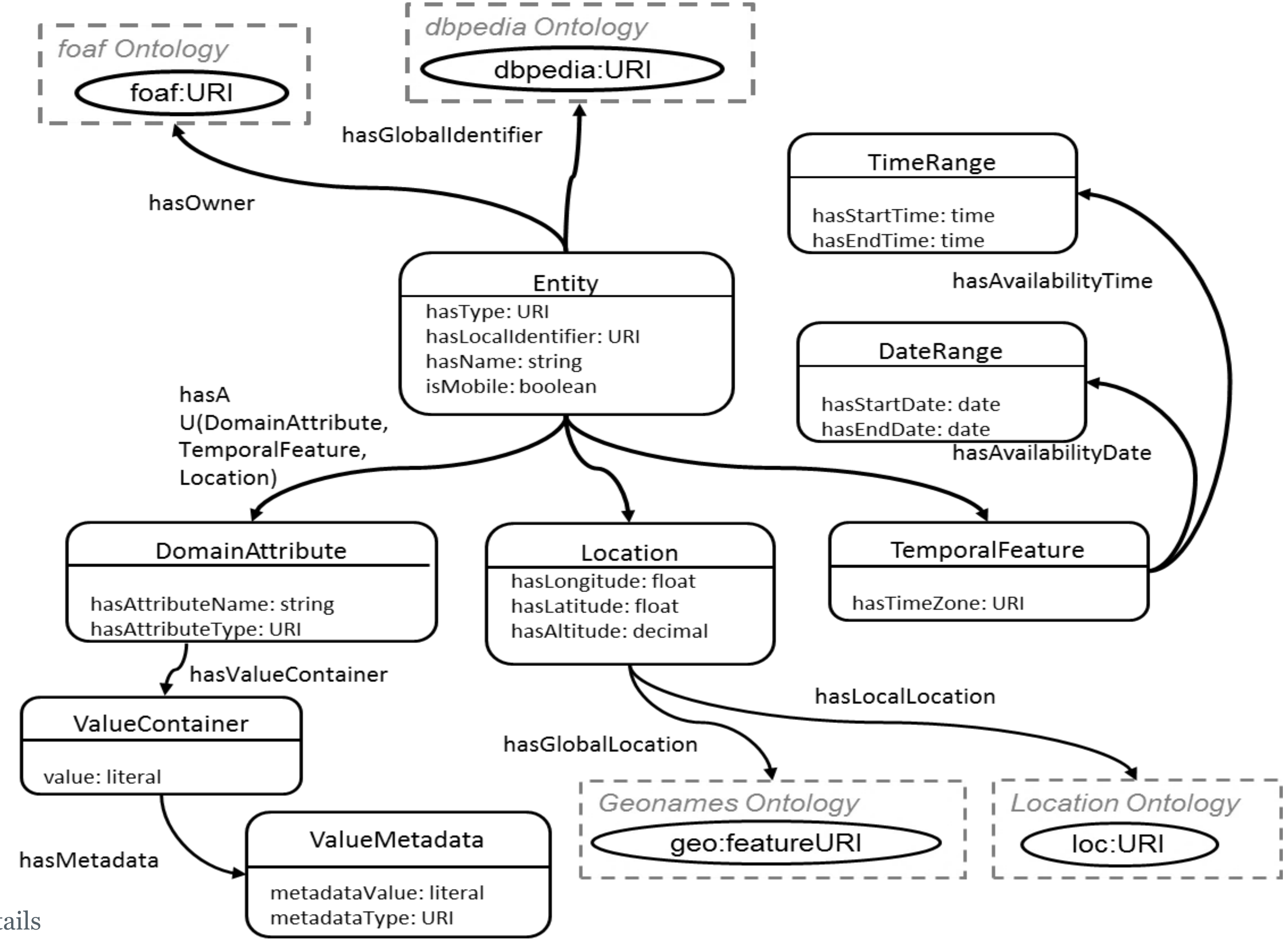
Role of Metadata

- Semantic tagging
- Machine-interpretable data annotation and resource descriptions
- Re-usable ontologies
- Resource description framework(s)
- Structured data, structured query

Metadata and Semantics

- To describe:
  - Content
  - Context
  - Resources
  - Entities and features of interest
- To create:
  - Perception
  - Situation awareness
- To support
  - Automated processes for management of resources and decision making

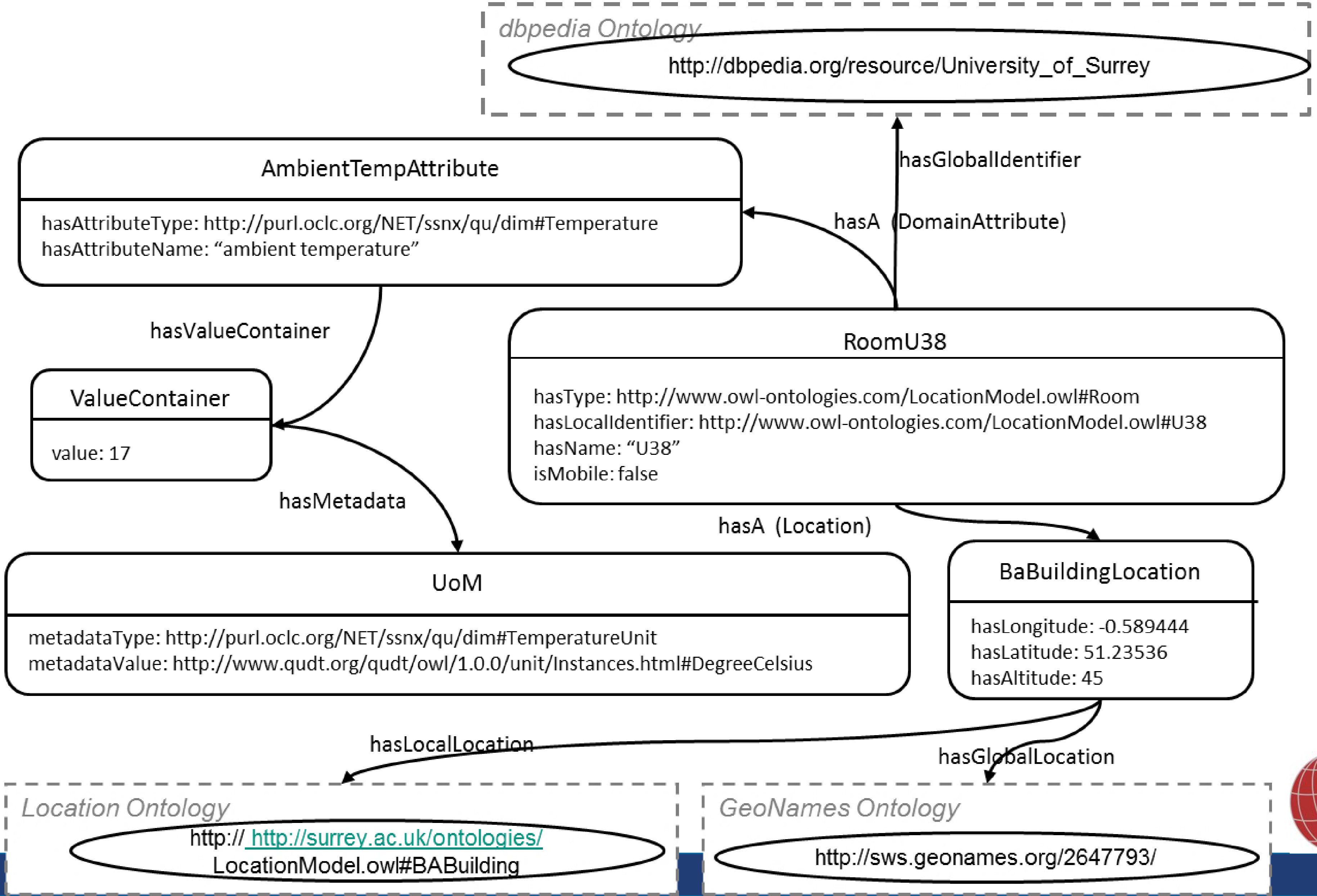
Entity Model



Refer to publication [1] for details

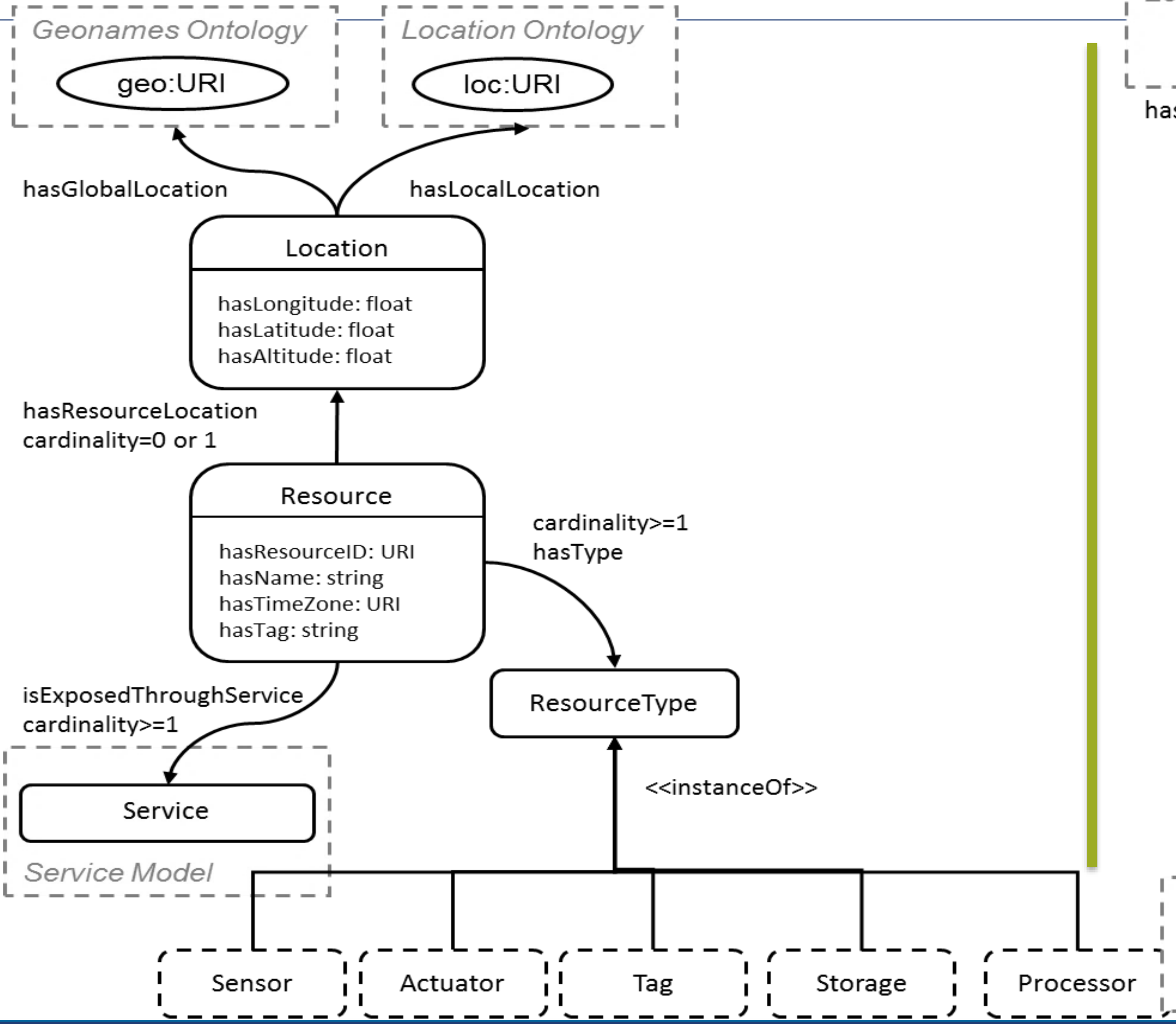
IoT-A
Internet of Things - Architecture

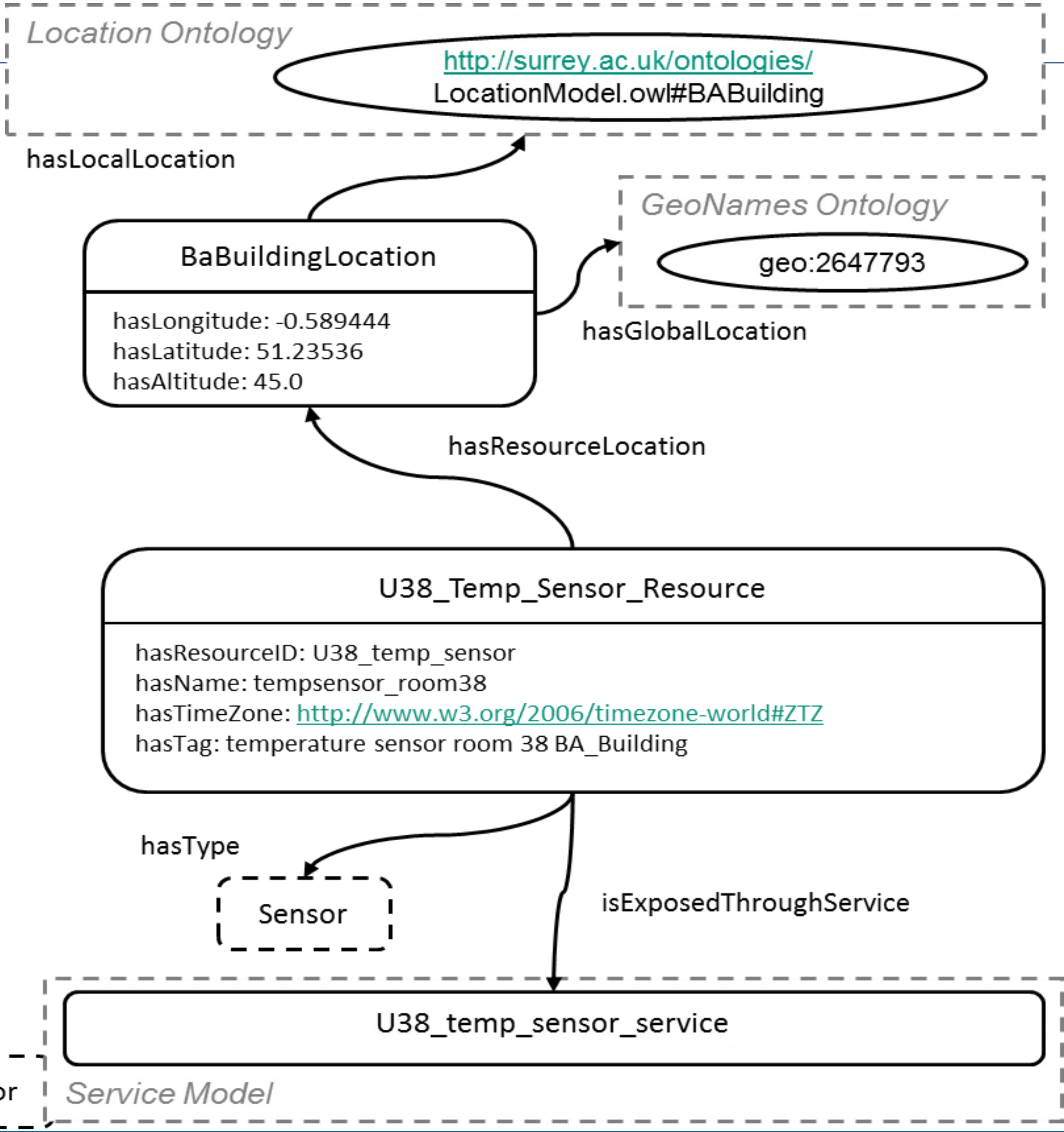
Entity Model Instance



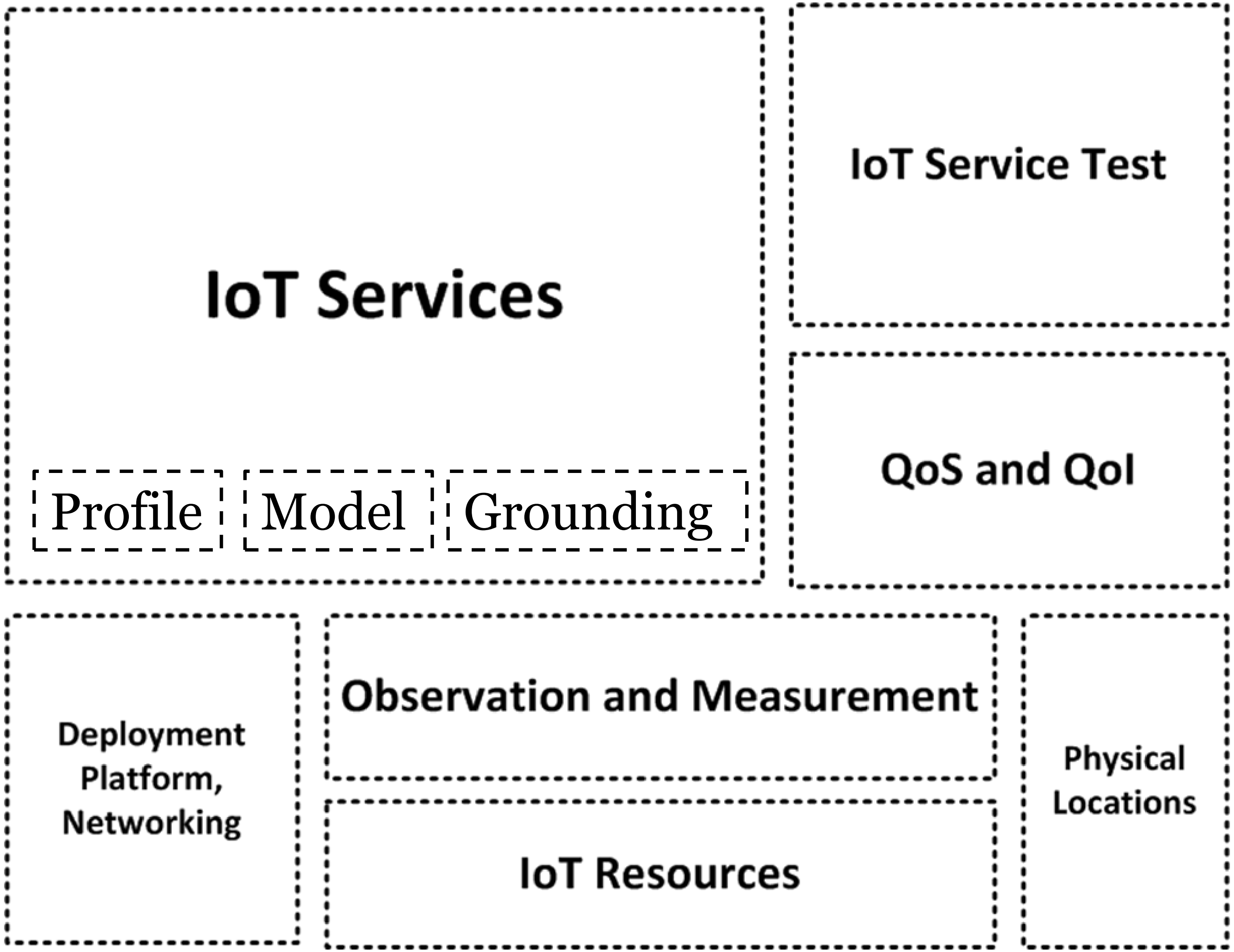
IoT-A
Internet of Things - Architecture

Resource Model and instance

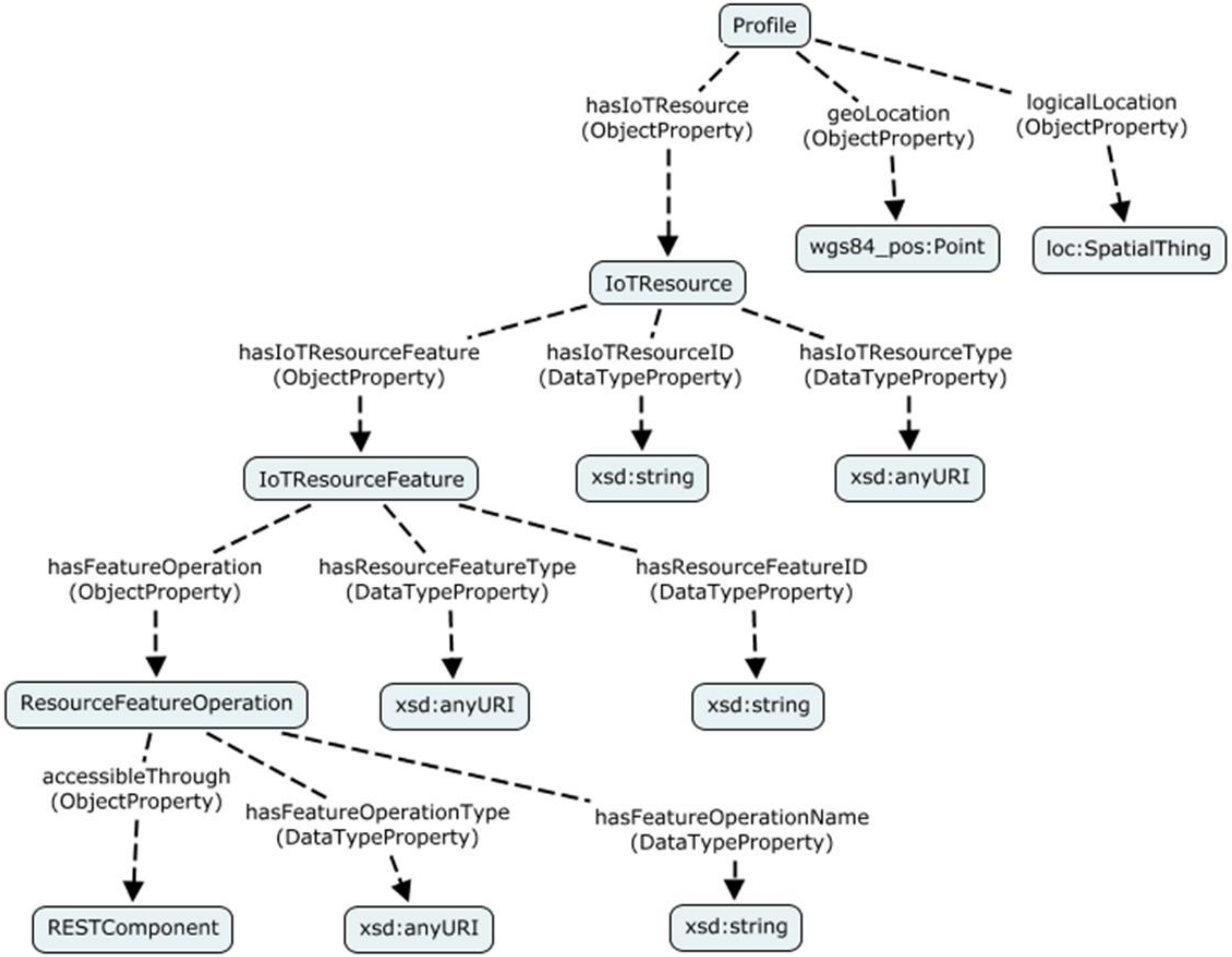




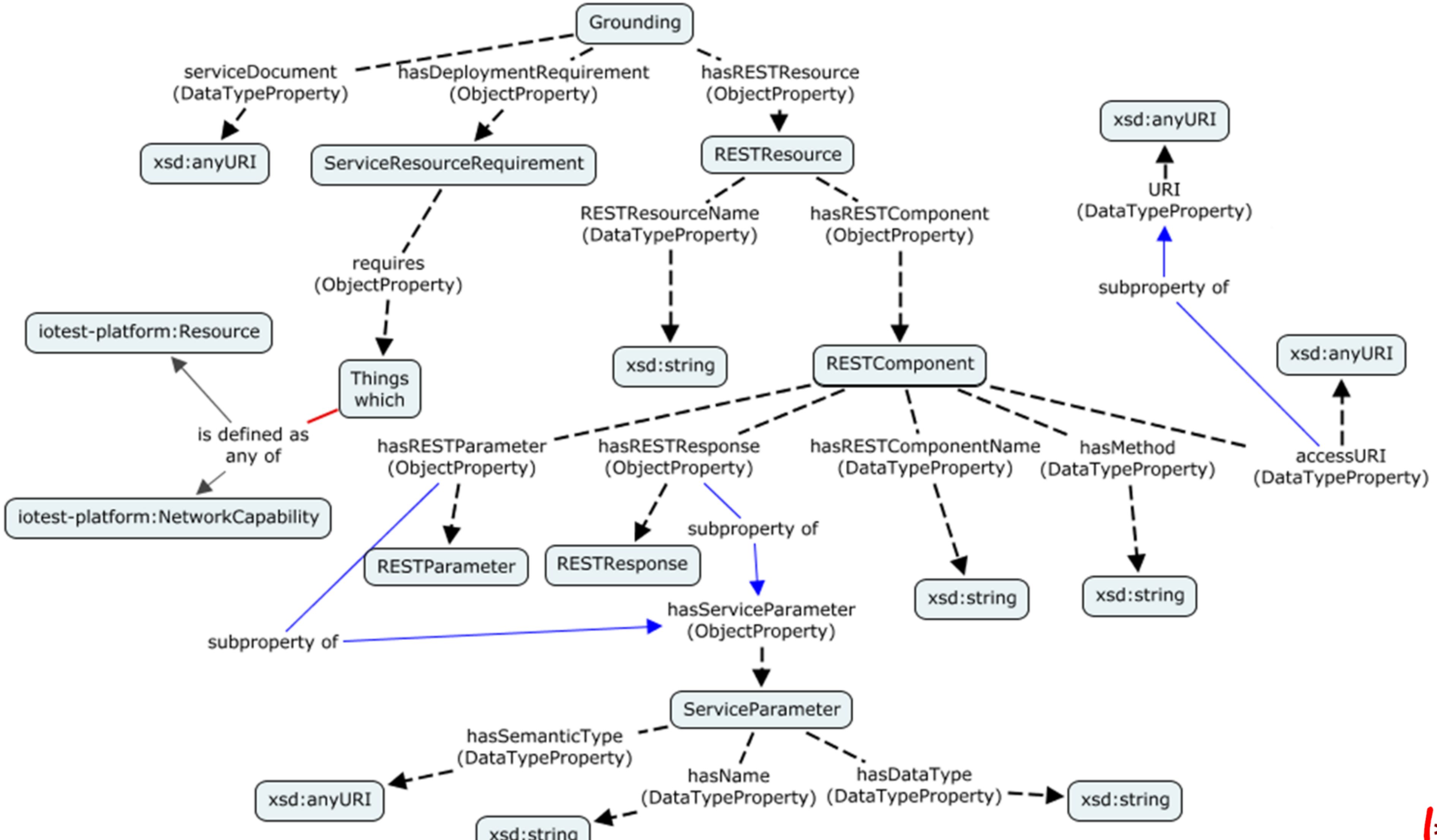
IoT Service model



IoT Service model: Profile

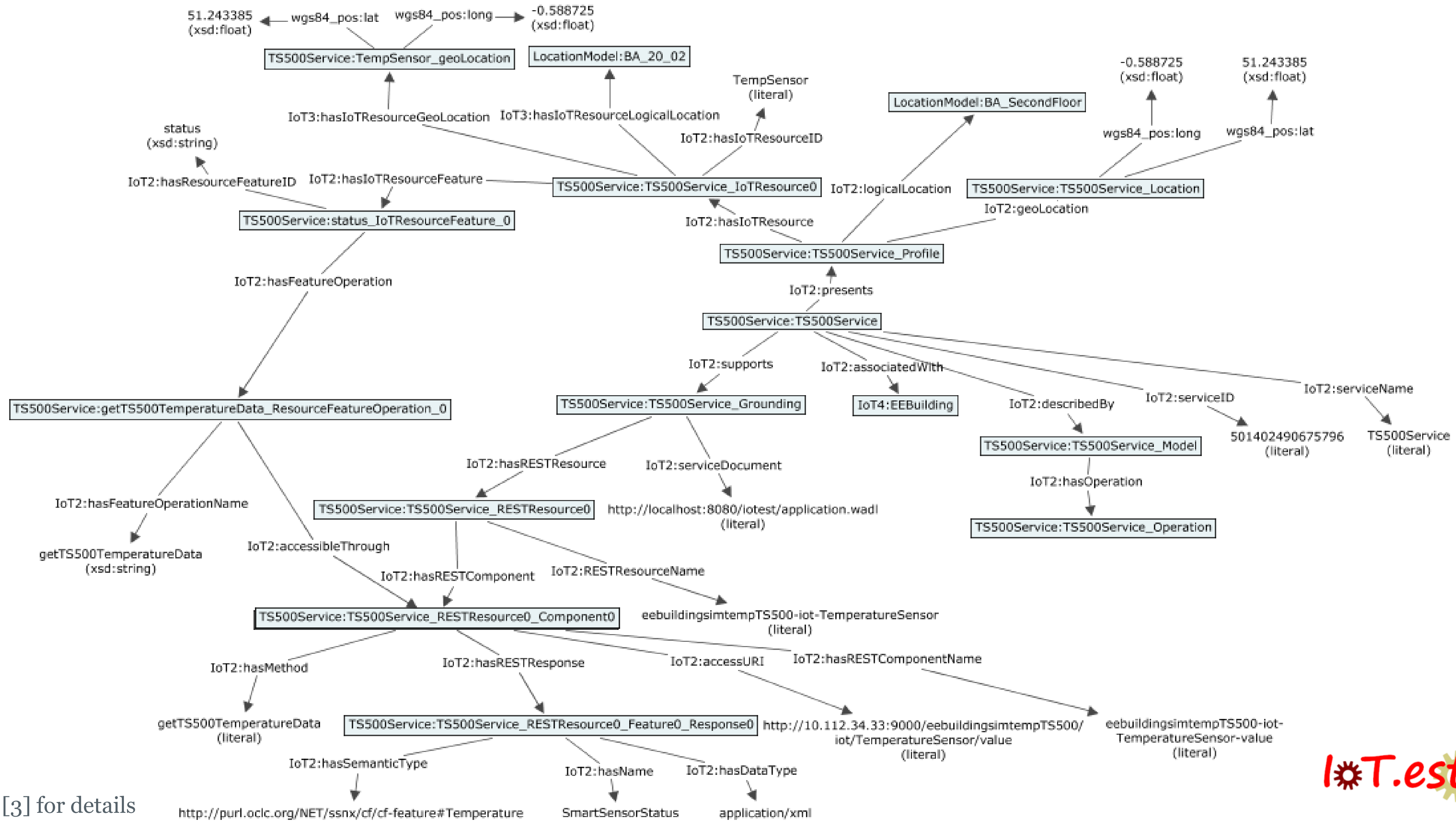


IoT Service model: Grounding



IoT.est

IoT Service model Instance

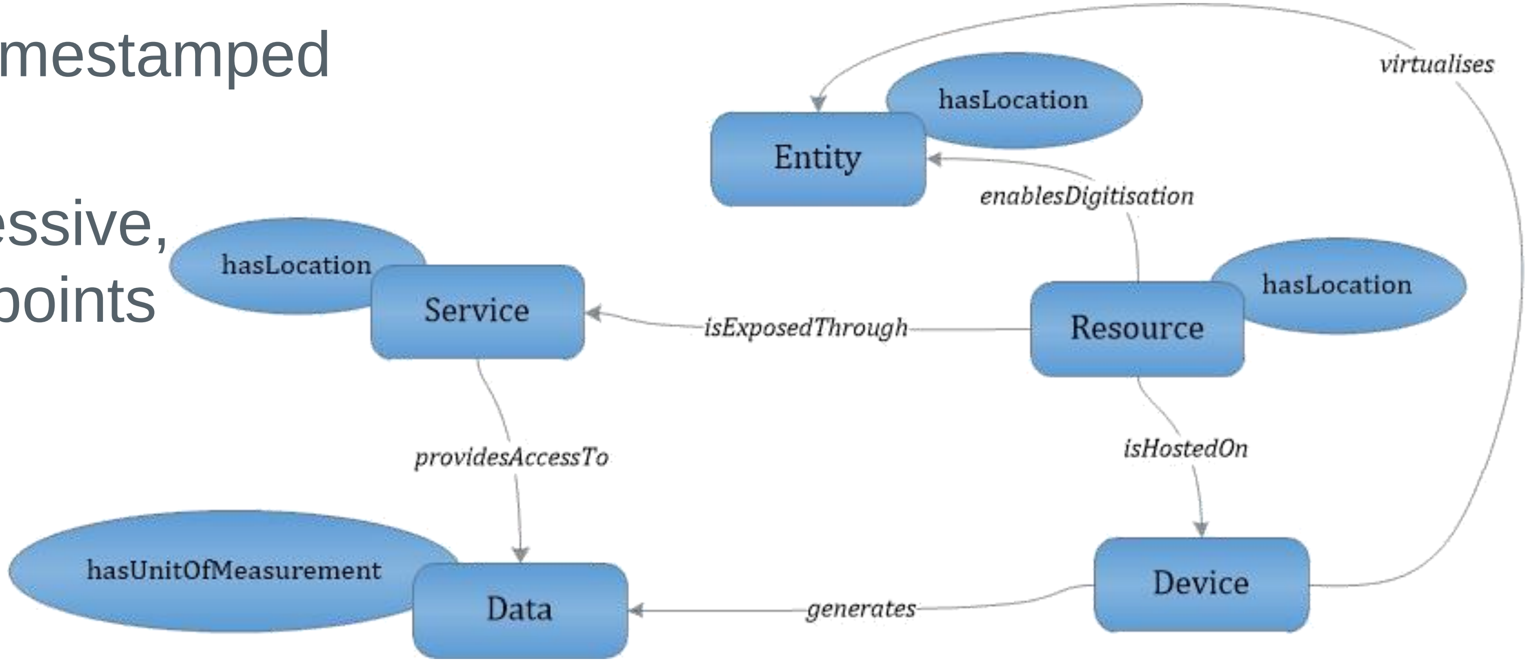


Refer to publication [3] for details

IoT.est

IoT Models: Summary

- Distinct repositories for metadata and data
- Metadata
  - Less frequency of update
- Data
  - frequently updated, timestamped and structured
  - not obtained at successive, equally-spaced time points

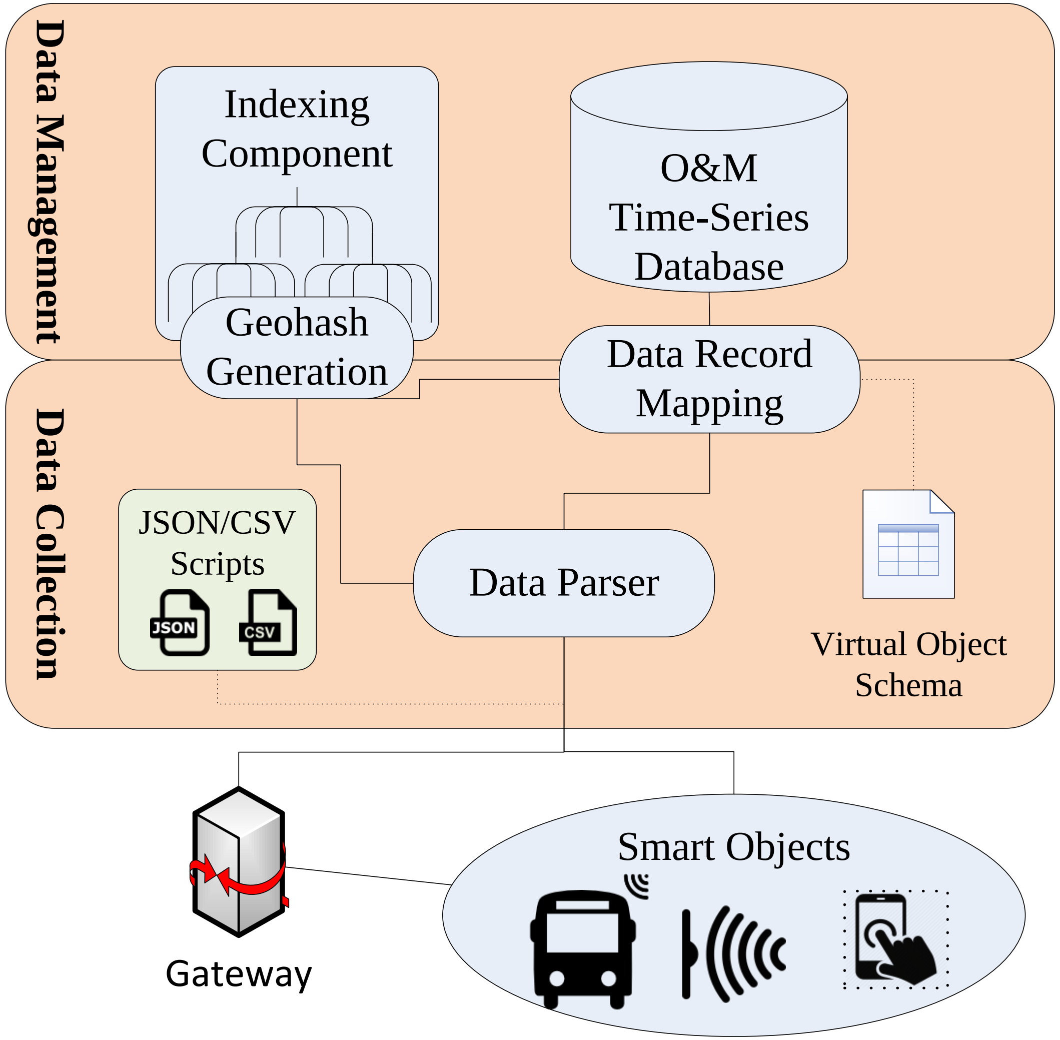


Refer to publication [4] for a survey of WoT ontologies

Data Storage and Retrieval

Data Indexing, Storage, and Retrieval Framework

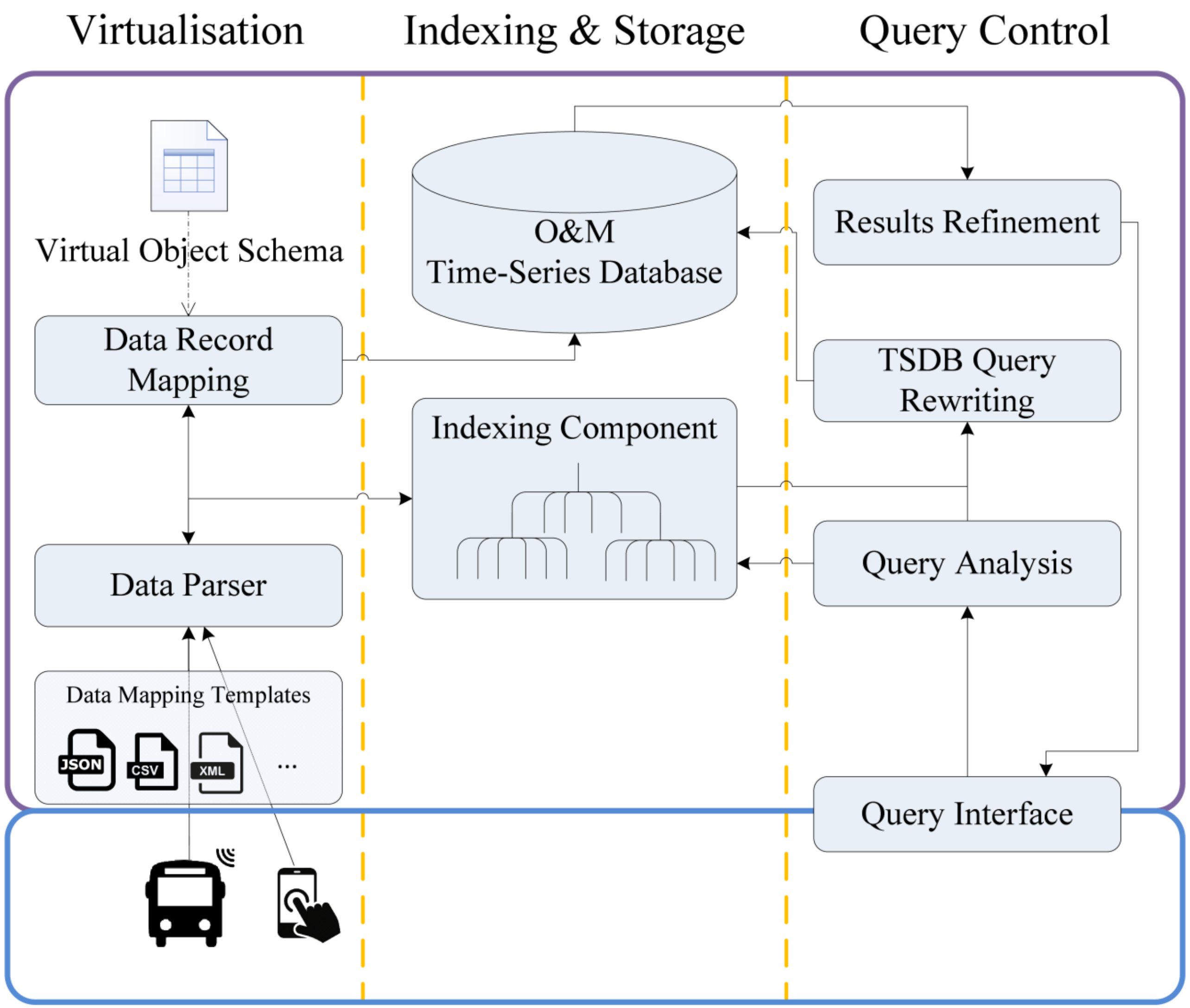
- Data Parser
  - Parse and transform from JSON, CSV, or other data
- Data Record Mapping
  - Map parsed data to format defined by Ontology Schema
- Spatial Indexing Component
  - Geohash-Grid Tree
  - Index spatial parameters of data records
- Time-Series Database (InfluxDB)
  - Store O&M data



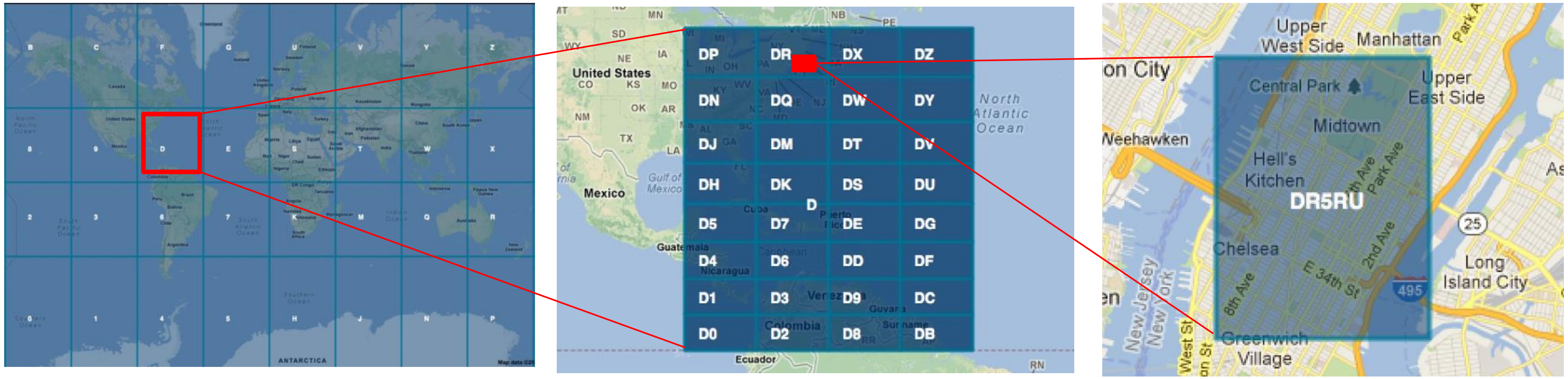
Refer to publication [6] for details.

Data Indexing, Storage, and Retrieval Framework

- Data Records
  - *a data record is a 5-tuple of the form [<object-id>, [<tag-key>=<tag-value>…(0..n)], [<field-key>=<field-value>…(1..n)], <geohash>, <unix-nano-timestamp>]*
- Considered Query
  - A query asking for observation and measurement (O&M) data based on spatial and temporal constraints



A Brief Overview of Geohash



Latitude and Longitude

– 40.8, -74.0

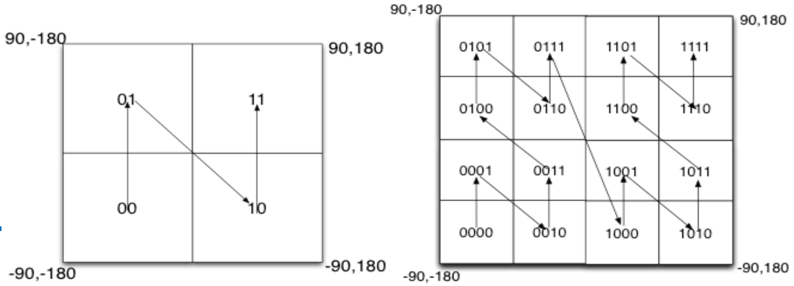
Binary/decimal representation

– 01100 10111 00101 10111

– 12 23 5 23 26

Geohash string

– dr5ru



Decimal	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Base 32	0	1	2	3	4	5	6	7	8	9	b	c	d	e	f	g

Decimal	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Base 32	h	j	k	m	n	p	q	r	s	t	u	v	w	x	y	z

[1] *Geohash. Available: http://en.wikipedia.org/wiki/Geohash*

[2] Paule. (2013). *Visualizing Geohash. Available: http://www.bigdatamodeling.org/2013/01/intuitive-geohash.html*

Data Parsing and Mapping

Indexing Component

eztpn45wn

Database	Measurement	Tag-key	Tag-value	Field-key	Field-value
mydb	vo_3021	geohash	eztpn45wn	*humidity*	0.64

InfluxDB Storage Mechanism

Parse

{ "**id**": "3021",
"**latitude**": "43.430007",
"**longitude**": "-3.949993",
"**title**": "bus3021",
"**image**": "http://lira.tlmat.unican.es/SmartSantander/iconos/tus.png",
"**content**": "<div class='googft-info-window'\n style='font-family: sans-serif; font-size: 10px;width: 200px; height: 18em ; overflow-y: auto;'><table width='100%' border='0'>\n <tr>\n <td valign='top'>
<h2 style='color: #5080e1'>NODE 3021</h2>Last update: 2015-01-02 17:33:19
Particles: 0.89 mg/m3
Humidity: 64.00%
</td>\n <td valign='top'></td>\n </tr>\n</table></div>",
"**tags**": "BUS"}
{...}

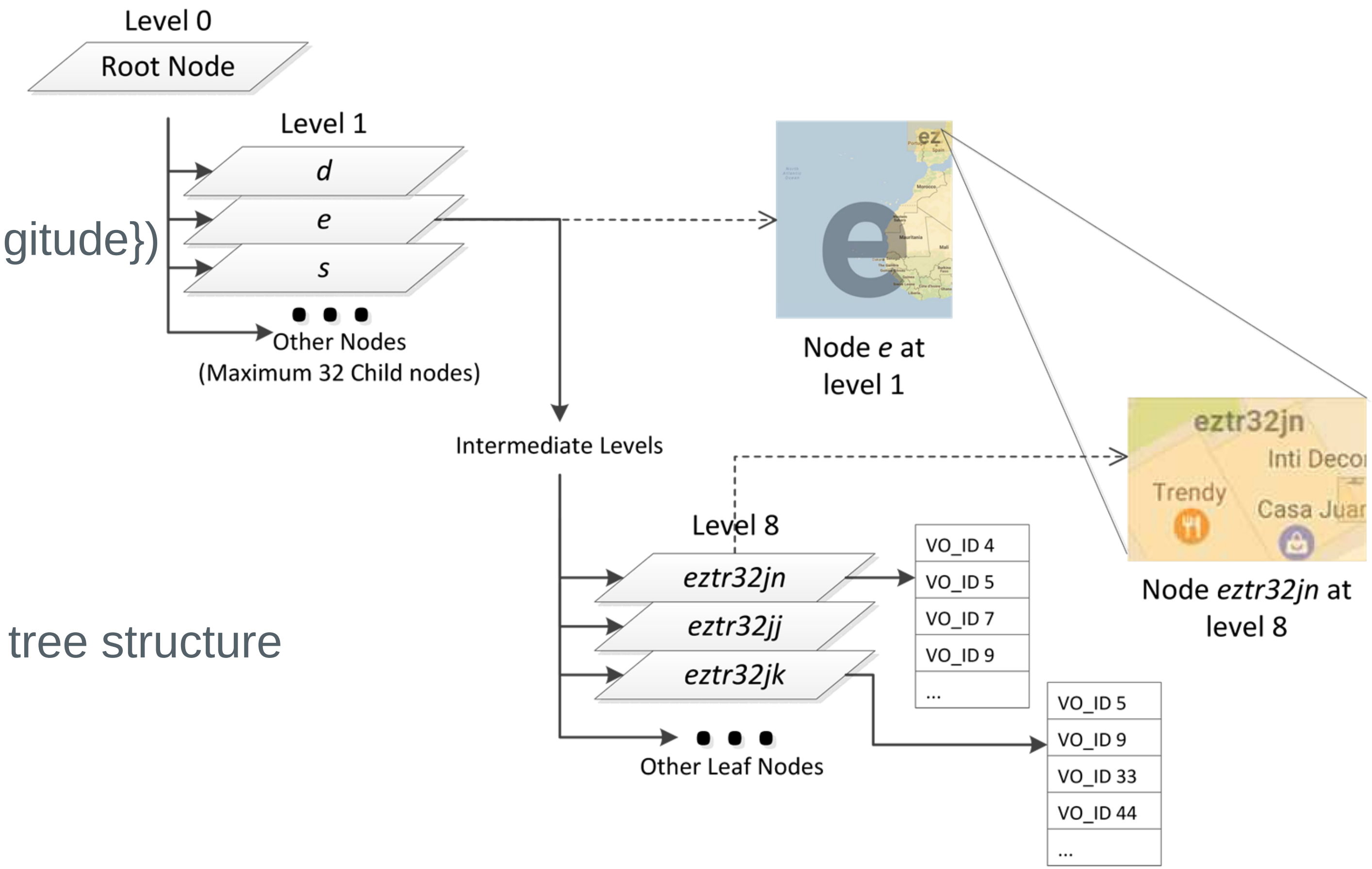
Raw data from SmartSantander (JSON)

mydb					
vo_3021					
geohash	*particles*	*humidity*	time	*latitude*	*longitude*
eztpn45wn	0.89	0.64	1420219999s	43.430007	-3.949993
…	…	…	…	…	…

InfluxDB

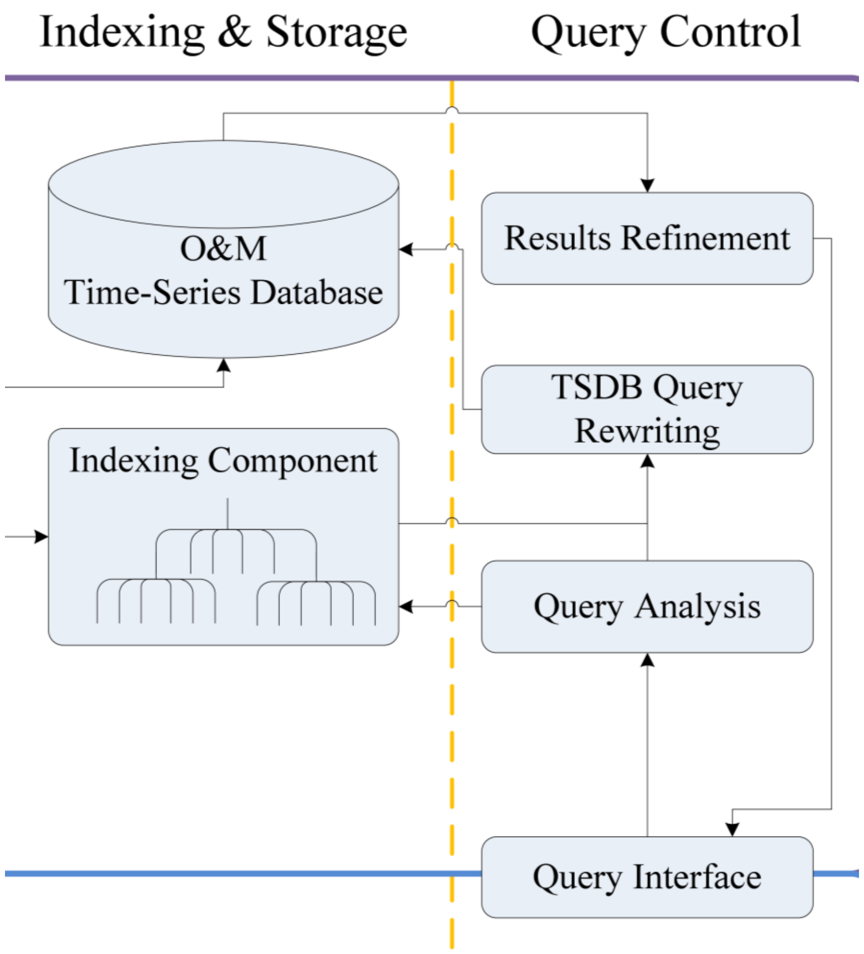
Spatial Index: Geohash-Grid Tree

- Node
  - Geohash string
  - Spatial grid (2 pairs of {latitude, longitude})
  - A list of stored VO_IDs (leaf node)
- Features
  - Unbalanced tree
  - Fixed height
  - Insertion without changing existing tree structure
  - No need to split node



Data Retrieval Components and Steps

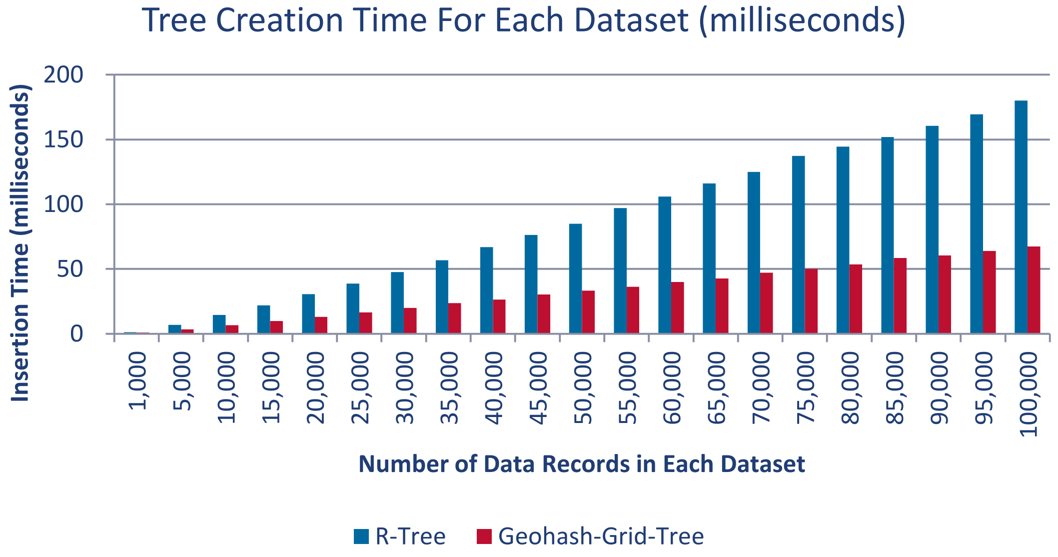
- Query Interface
  - Get query
- Query Analysis
  - Analyze query
  - Send request with spatial constraints to indexing component to get matched IDs
- Query Rewriting
  - Rewrite query with matched IDs and other constraints
- Results Refinement
  - Refine returned data from database into a proper format for display on query interface

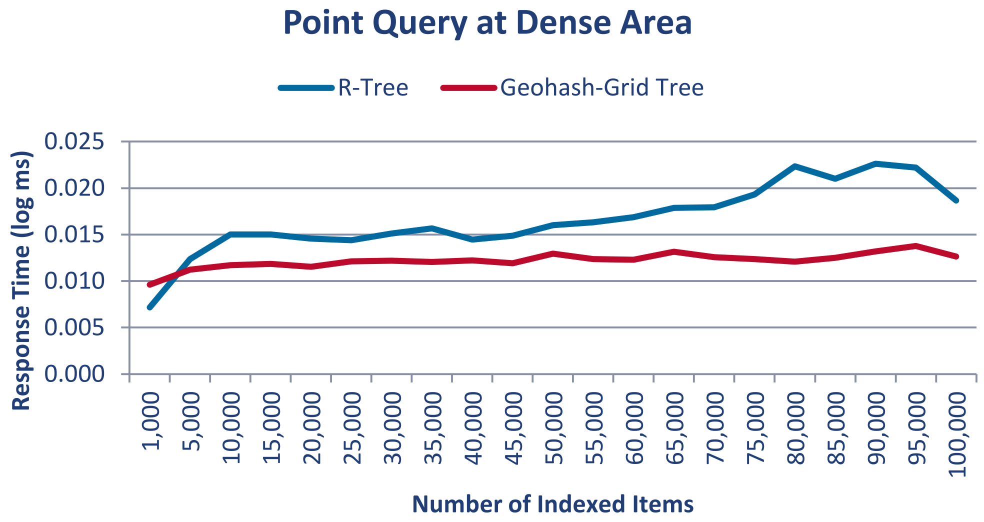


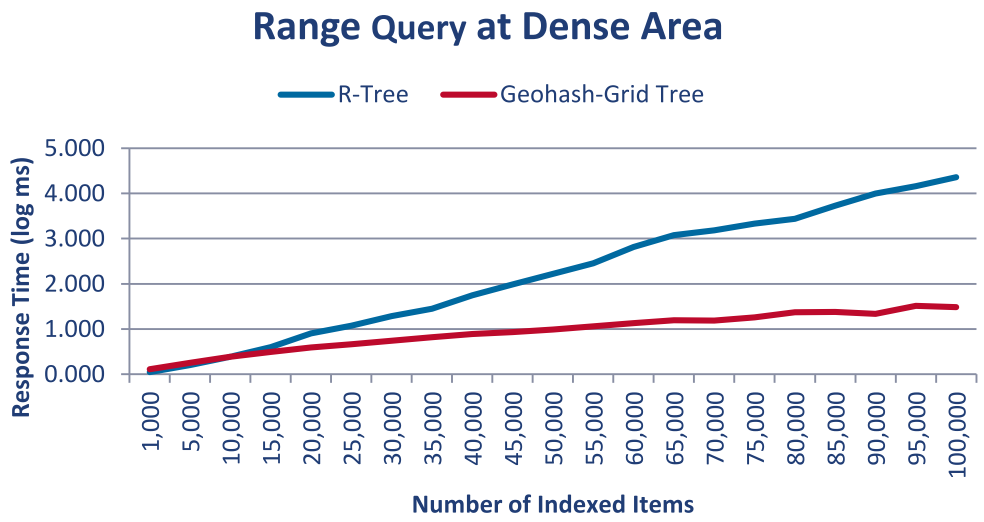
Spatial Index: Performance measurements

Geohash-Grid Tree Comparing to R Tree

- Insertion does not change existing tree structure
- Fast indexing creation time
- Better query response time in dense areas

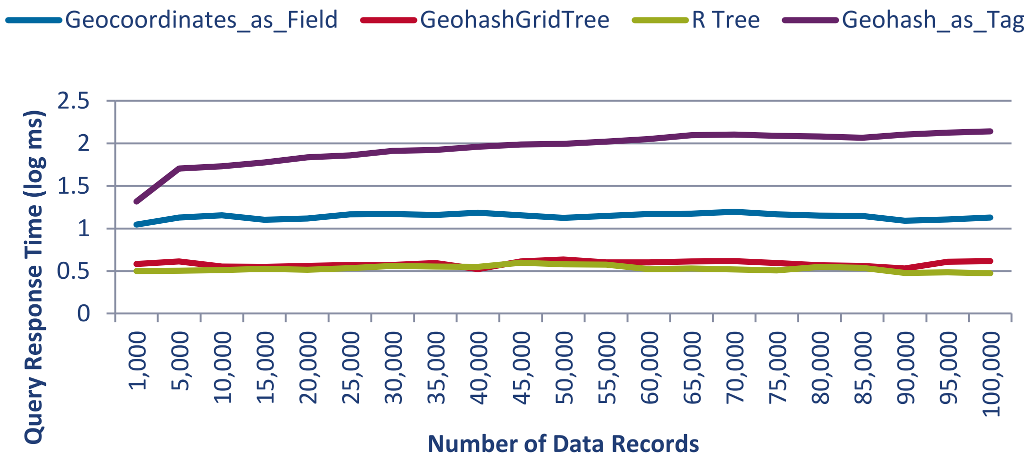


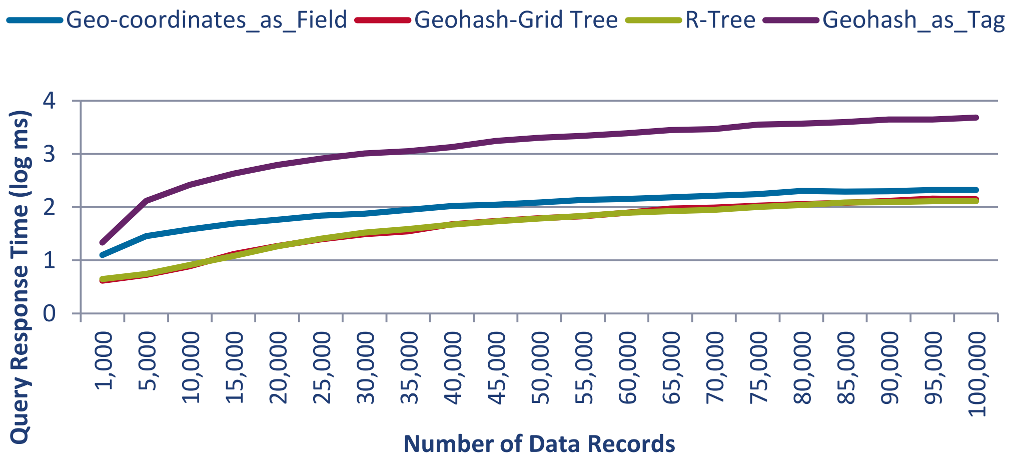




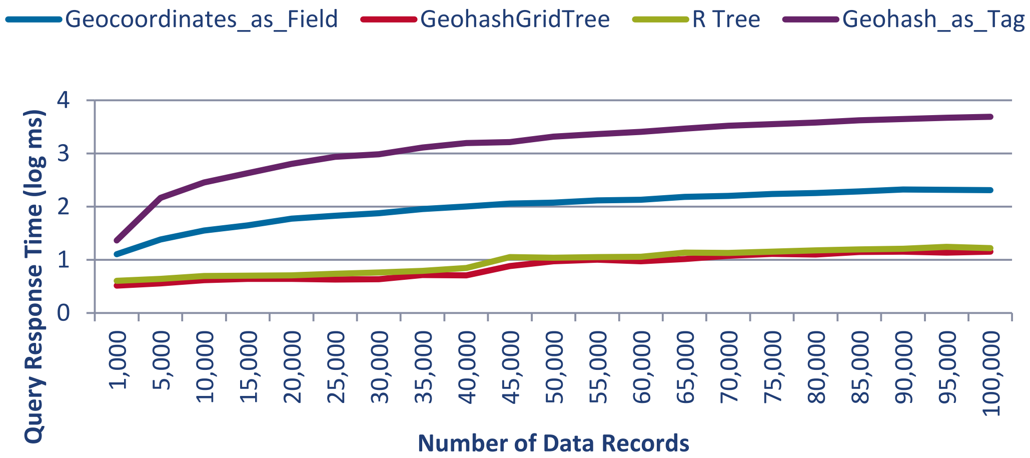
Spatial Indexing and Retrieval: Performance measurements

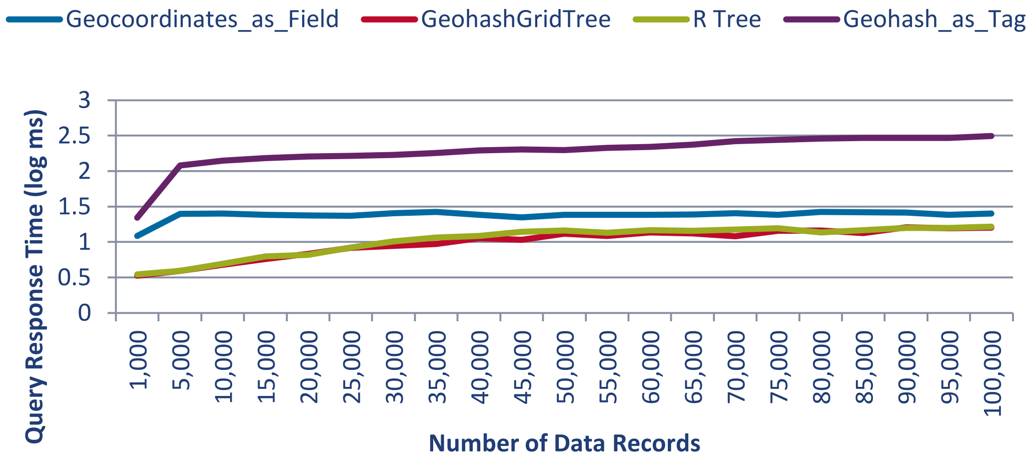
Short time period





Long time period





Sparse area

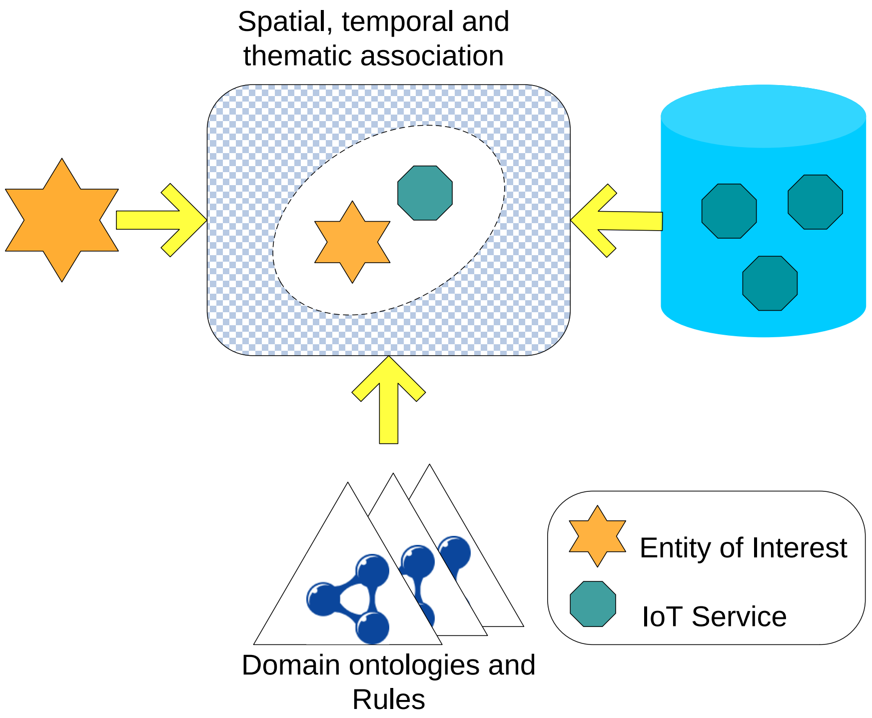
Dense area

Data Analysis

Case Study I: Smart Campus – IoT testbed

Semantic Reasoning for Association Analysis

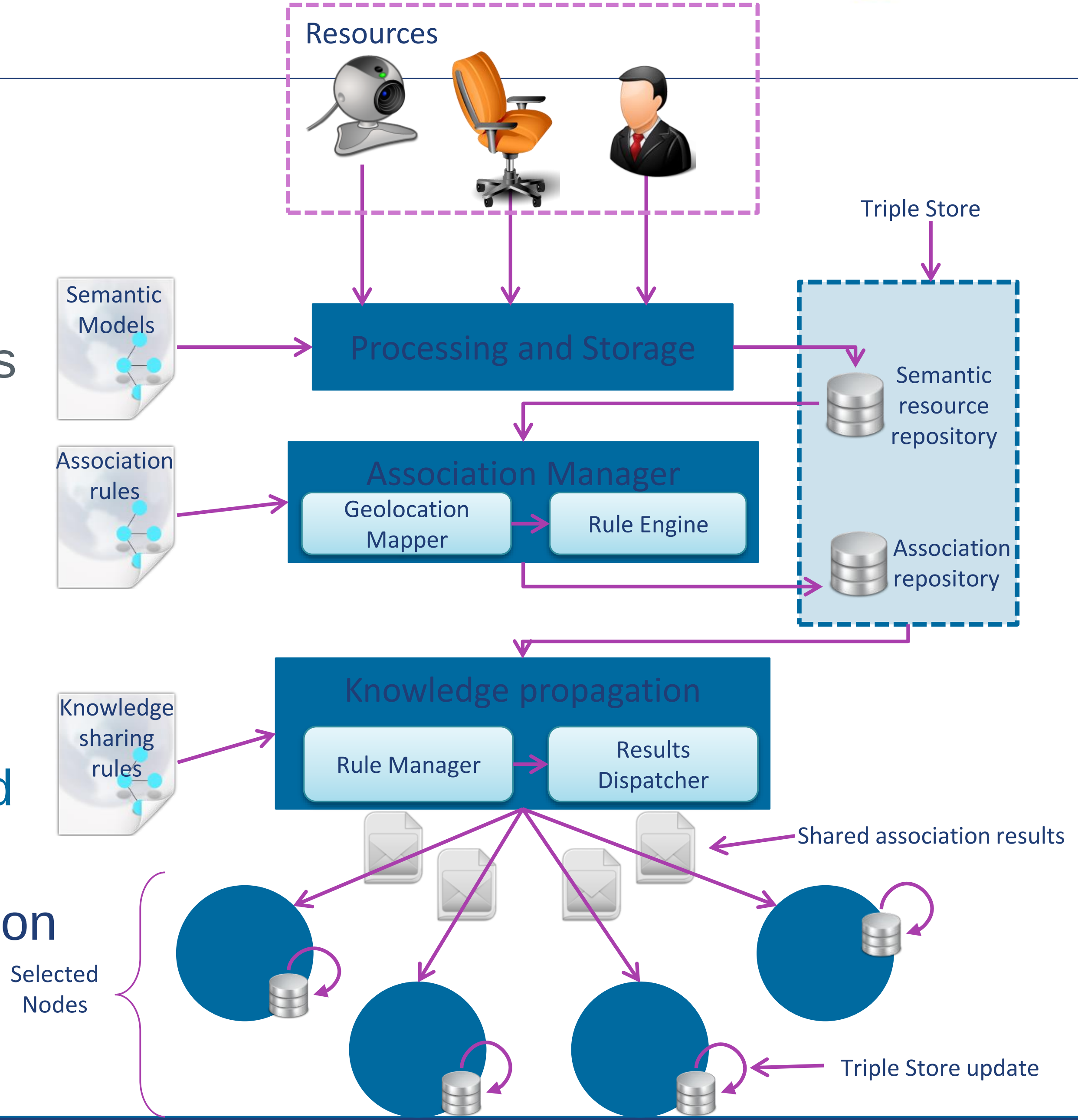
- Associations along thematic-spatial-temporal axes
  - Thematic (feature) match – utilising domain ontologies that capture virtual entity's attributes and IoT Service's input/output parameter
  - Spatial match – utilising location ontologies that model logical locations with properties such as 'contains'
  - Temporal match – utilising temporal aspects of entities which have a temporal aspect, such as meeting rooms with the IoT Service's observation_schedule
- Dynamic association inference through
  - Rules that incrementally reason on feature, spatial aspects and time
- Provision for semantic queries on derived associations



Refer to publication [2] for details.

Smart Campus

- Driven with semantic models describing
  - Entities
    - Campus buildings, floors and rooms
  - Location model capturing indoor locations
    - Rooms, corridoors etc. with their proximity, containment relationships
  - Resources
    - Temperature, light sensors
  - IoT services
    - Access interface to IoT resources and their O&M data
- Dynamic thematic-spatial-temporal association inference



IoT-A Internet of Things - Architecture

Refer to publication [2] for details.

Smart Campus: SWRL rules for thematic-spatial associations

Rule-1:

srv:Service(?s) ∧ srv:hasOutput(?s, ?out) ∧ em:Entity(?et) ∧ em:hasA(?et, ?da) ∧ em:hasAttributeType(?da, ?atype) ˚ sqwrl:makeSet(?sr, ?out) ∧ sqwrl:groupBy(?sr, ?s) ∧ sqwrl:makeSet(?se, ?atype) ∧ sqwrl:groupBy(?se, ?et) ˚ sqwrl:intersection(?in, ?sr, ?se) ∧ sqwrl:size(?n, ?in) ∧ swrlb:greaterThan(?n, 0) → assoc:sameFeatureAs(?s, ?et)

Rule-2:

assoc:sameFeatureAs(?s, ?et) ∧ srv:hasServiceArea(?s, ?sa) ∧ em:Entity(?et) ∧ em:hasA(?et, ?l) ∧ em:hasLocalLocation(?l, ?loc) ˚ sqwrl:makeSet(?rsa, ?sa) ∧ sqwrl:groupBy(?rsa, ?s) ∧ sqwrl:makeSet(?eloc, ?loc) ∧ sqwrl:groupBy(?eloc, ?et) ˚
sqwrl:intersection(?in, ?rsa, ?eloc) ∧ sqwrl:size(?n, ?in) ∧ swrlb:greaterThan(?n, 0) → assoc:isAssociatedWith(?s, ?et)

Rule-3:

assoc:sameFeatureAs(?s, ?et) ∧ srv:hasServiceArea(?s, ?sa) ∧ em:Entity(?et) ∧ em:hasA(?et, ?l) ∧ em:hasLocalLocation(?l, ?loc) ∧ loc:givesAccessTo(?sa, ?loc) → assoc:isAssociatedWith(?s, ?et)

Rule-4:

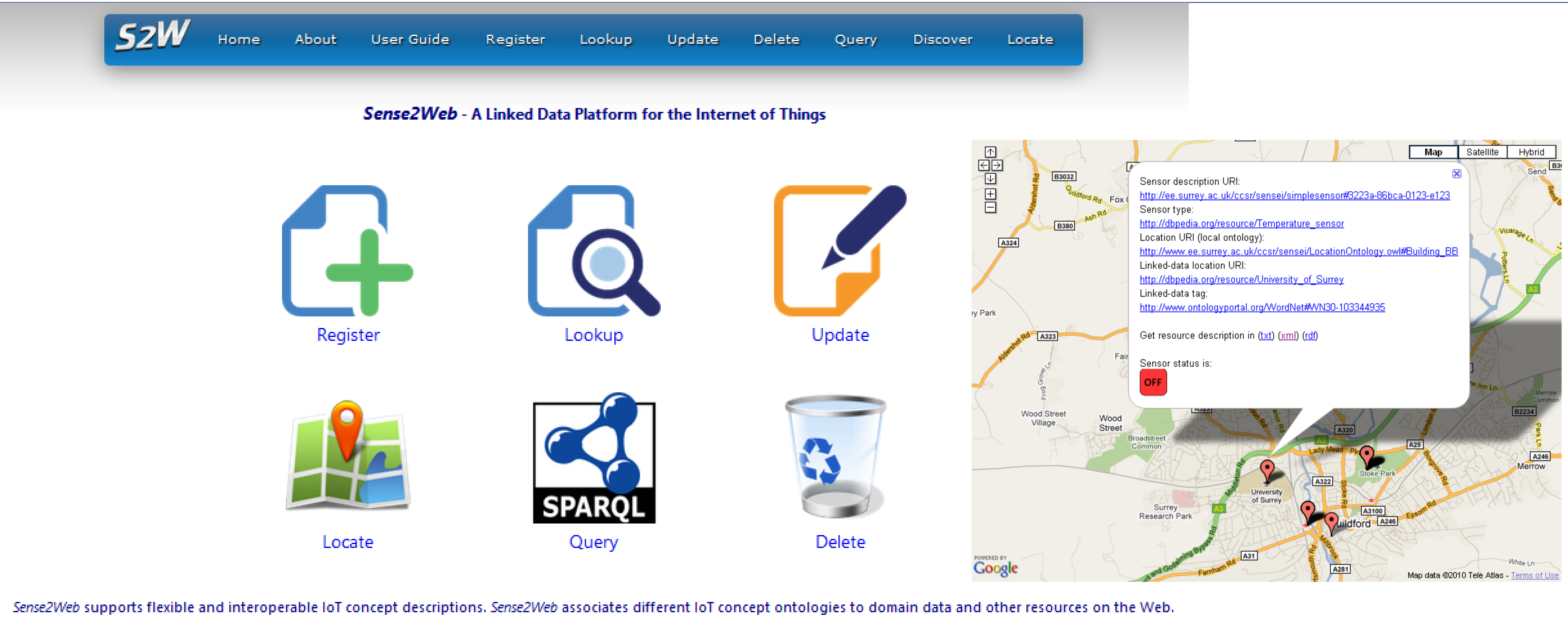
assoc:sameFeatureAs(?s, ?et) ∧ srv:hasServiceArea(?s, ?sa) ∧ em:Entity(?et) ∧ em:hasA(?et, ?l) ∧ em:hasLocalLocation(?l, ?loc) ∧ loc:isAdjacentTo(?sa, ?loc) → assoc:isAssociatedWith(?s, ?et)

Property name Description	Domain	Range
Contains	Place	Place
Allows a place to contain other places (e.g. a floor containing some rooms)		
isAdjacentTo	Place	Place
Models that two places are separated by some boundaries		
inEast	Place	Place
inWest	Place	Place
inNorth	Place	Place
inSouth	Place	Place
Refinement of isAdjacentTo, including the cardinal direction(s) of a place relatively to another		
givesAccessTo	Place	Place
Means that a door exists in the boundary separating two places connecting them		
isIncludedIn	Place	Place
Inverse property of 'contains'		
isPrivate/isPublic/isSemiPrivate	Place	Boolean
Allows to know if a place can be used or not when computing associations		

IoT-A
Internet of Things - Architecture

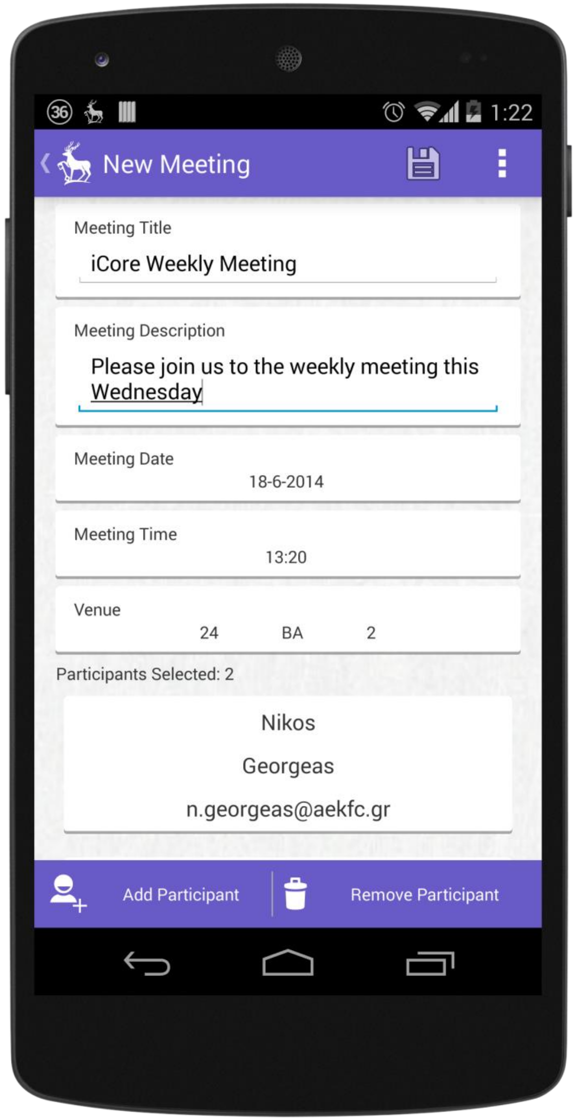
Refer to publication [2] for details.

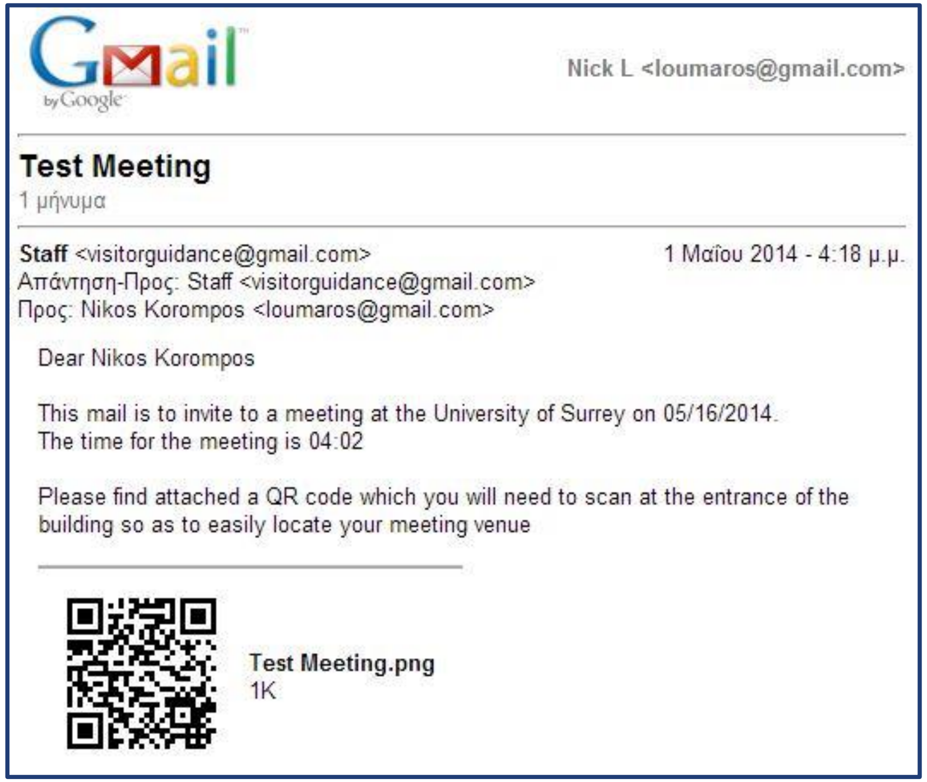
Smart Campus: Online platform

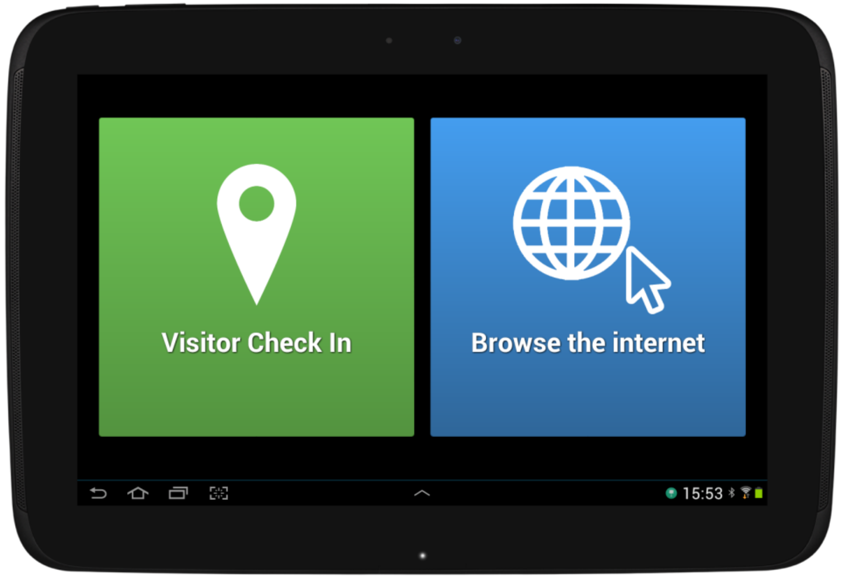


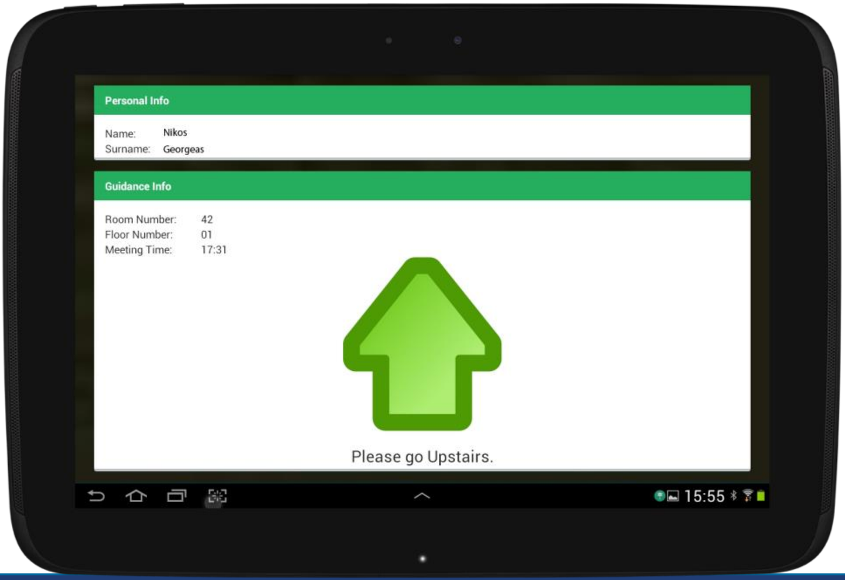
Ref: T. Elsaleh *et al*. Sense2Web Linked Data Platform; http://iot.ee.surrey.ac.uk/s2w/. Details in reference [1]

Smart Campus: Meeting venue guidance









iCore
Empowering IoT through Cognitive Technologies

Case Study II: Recycling FMCG

Semantic modelling of Smart tags

- Smart tags for:
- Ice cream packaging
  - QR code
    - Unique identifier for each ice cream
  - Time temperature indicator (TTI) label
    - indicator of the quality of the ice cream based on its cold chain
    - irreversible thermochromic ink that will change its colour after exposure to a temperature above -10 C° for more than 30 minutes

Semantic modelling of Smart tags

- Representation of both
  - printed electronics (e.g. NFC) and
  - passive printed 2D tags such as datamatrix and QR codes (encoded using dynamic inks)
    - that can be read by scanners
- Semantic model for SmartTags capturing:
  - reactions to chemical/physical conditions [e.g. Inks could be thermochromic, hydrochromic or fluorescent visible/invisible, i.e. they ‘reactTo’ temperature/humidity/light within defined ranges]
  - reaction state (reversible/permanent)
  - status (activated/expired/not-expired)
  - links to required decoding mechanism
  - links to recorded measurements, including context data (location, time etc.)

TAG IT SMART

Semantic Analysis for FMCG

- System provides point-of-recycling information for every consumer packaged good (CPG)
- Allows tracking of the FMCG lifecycle
- Enables cold chain visible quality indicator
- Semantic modelling can enable:
  - Detection of erroneous QR code scan measurements
  - Generic, enhanced recommendations on (nearest and relevant) recycling points

Unilever dondelotiro.com EVRYTHNG THINFILM Memory Everywhere TAG IT SMART

Case Study III: Cyber-Physical-Social (CPS) data analysis, fusion

CPS Data Fusion

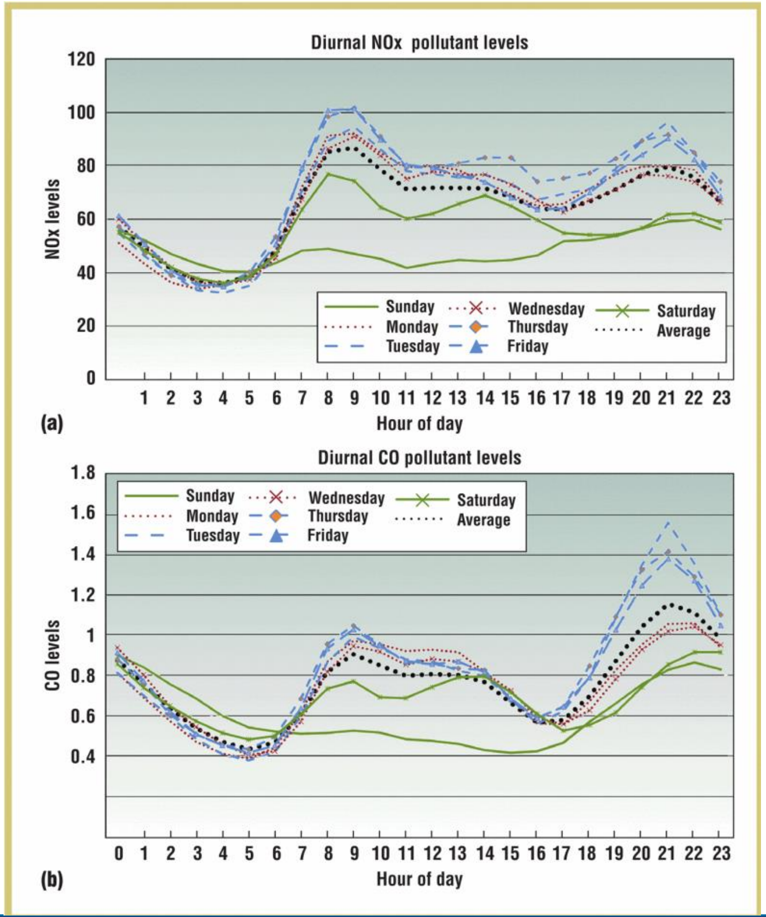
- Social data source:
  - Recorded Foursquare check-ins in Patras, Greece for 3 months
    - 100 Days between July to September 2012
    - Created a grid of “listening posts” that sampled foursquare API every 30 minutes
    - Each listening post queries API to retrieve names of businesses within their range, current check-ins and total check-ins. From this data the number of users who checked-in within the last 30 minutes can be calculated
    - 282 venues recorded of which 249 checked into
    - Average of 145.82 Check-ins per day

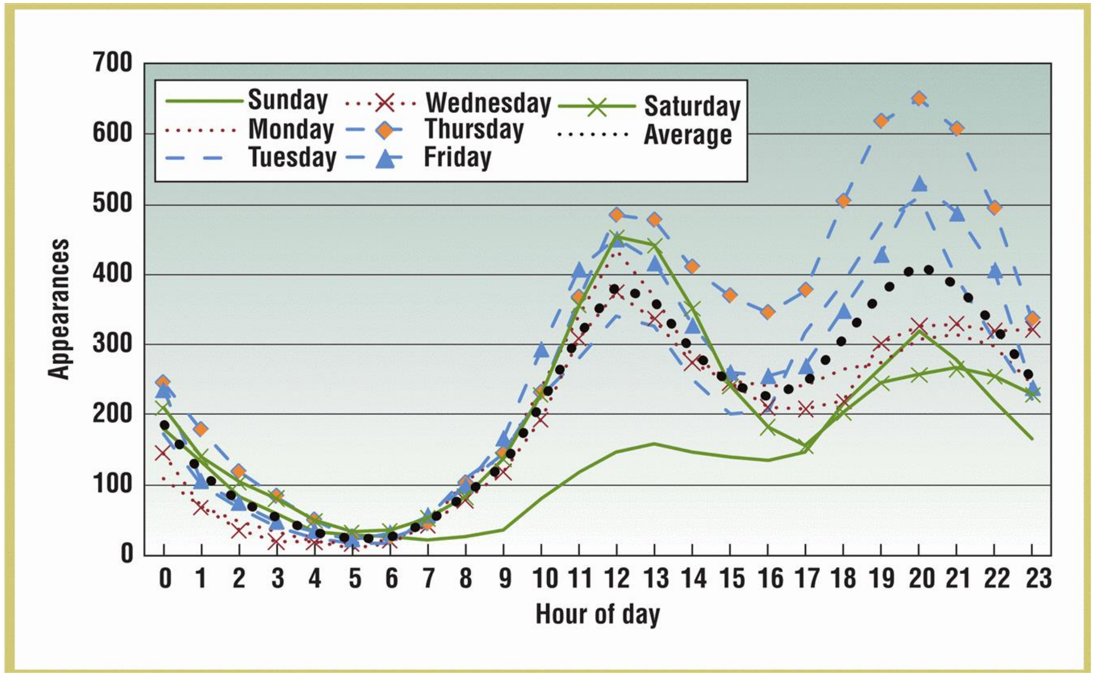
- Physical world data: Traffic and Air Quality
  - Network of 29 stations where measurements taken
  - Blue stations where traffic measurements taken over single 24-hour period, Pink stations where traffic measurements taken over 7 day period
  - X marks the air quality monitoring station.
  - Air pollution measurements provided by the public repository of the Hellenic Ministry of Environment, Energy and Climate Change

Refer to publication [7] for details.

CPS Data Fusion

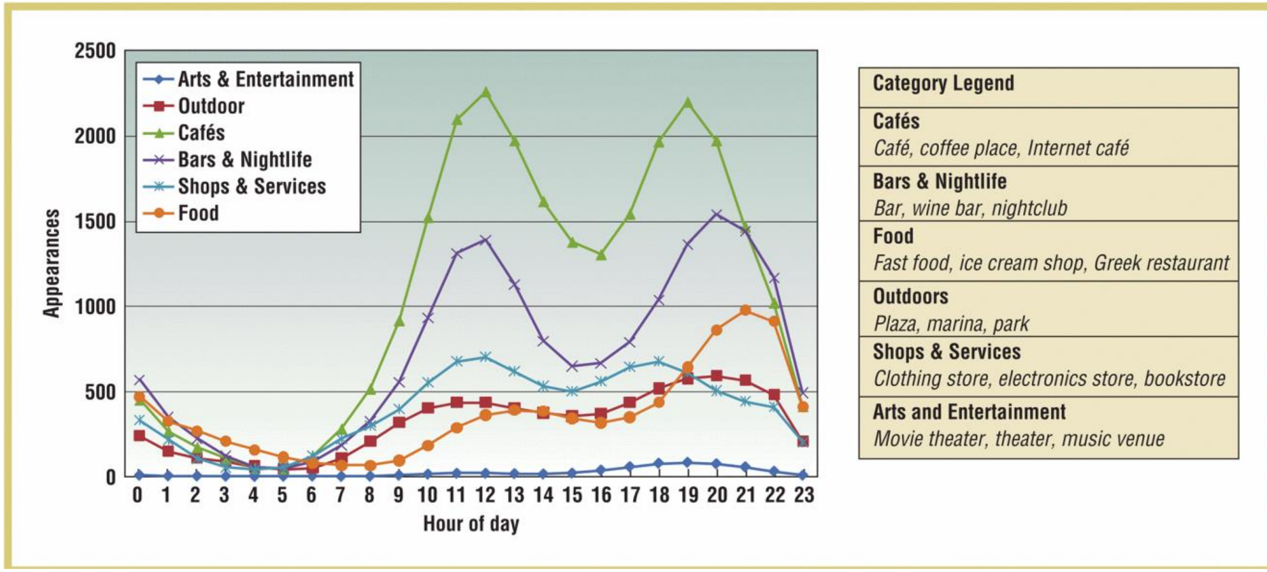
- Correlated Air Quality (NO and CO) and Traffic data during a single 24 hour period (2 stations were over 1 week) with Foursquare check-in data
- Correlation found between diurnal Foursquare check-ins, traffic volume, NOx and CO pollutants





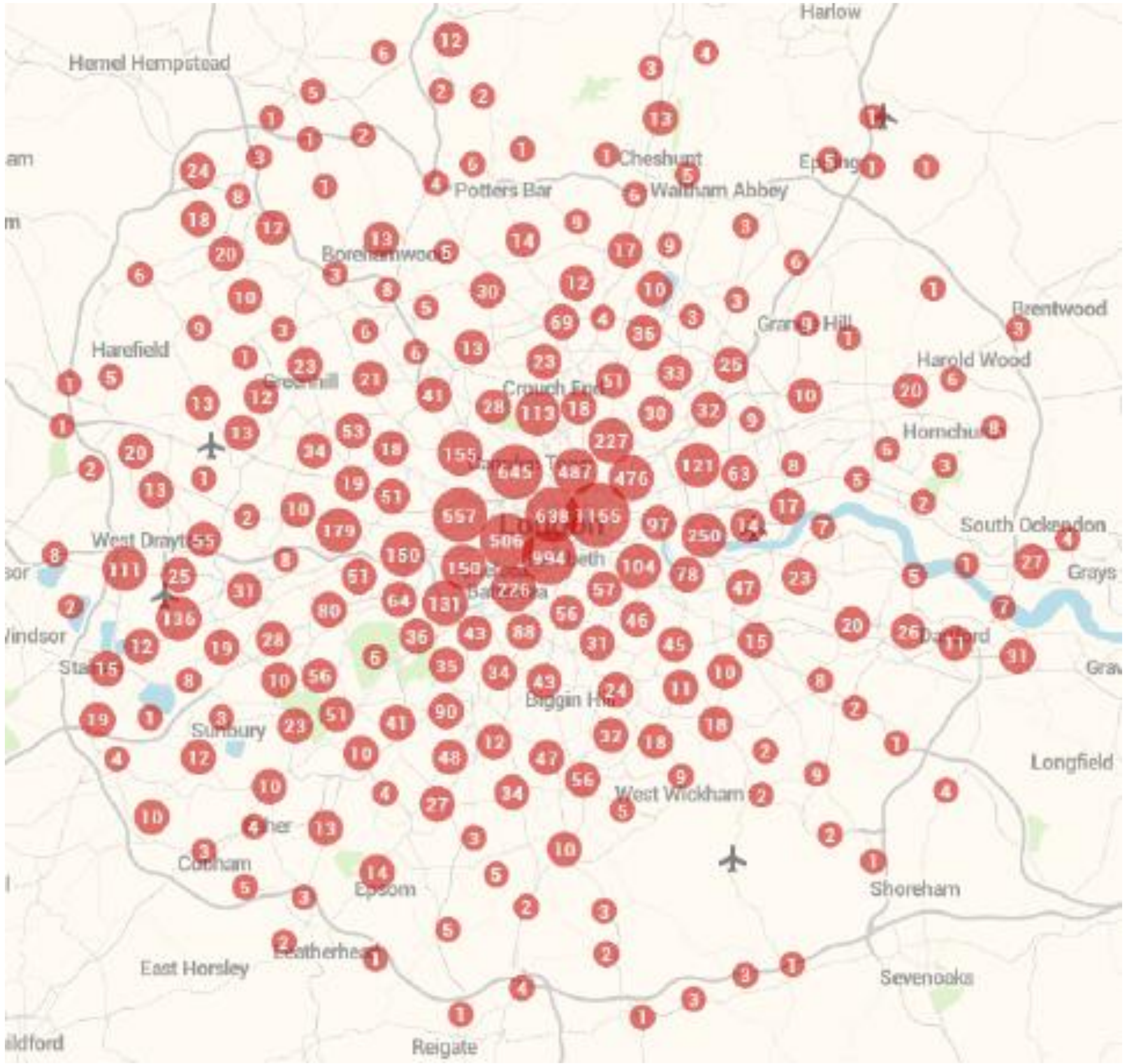
CPS Data Fusion

- Multilevel categories of check-ins used to better understand peoples diurnal patterns
  - was used to compare activity popularity, activity times and compare venues
  - only a few venues take up the majority of checkins (Top 20% of venues take up 69.2% of checkins).



CPS Data Analysis: City Rhythms

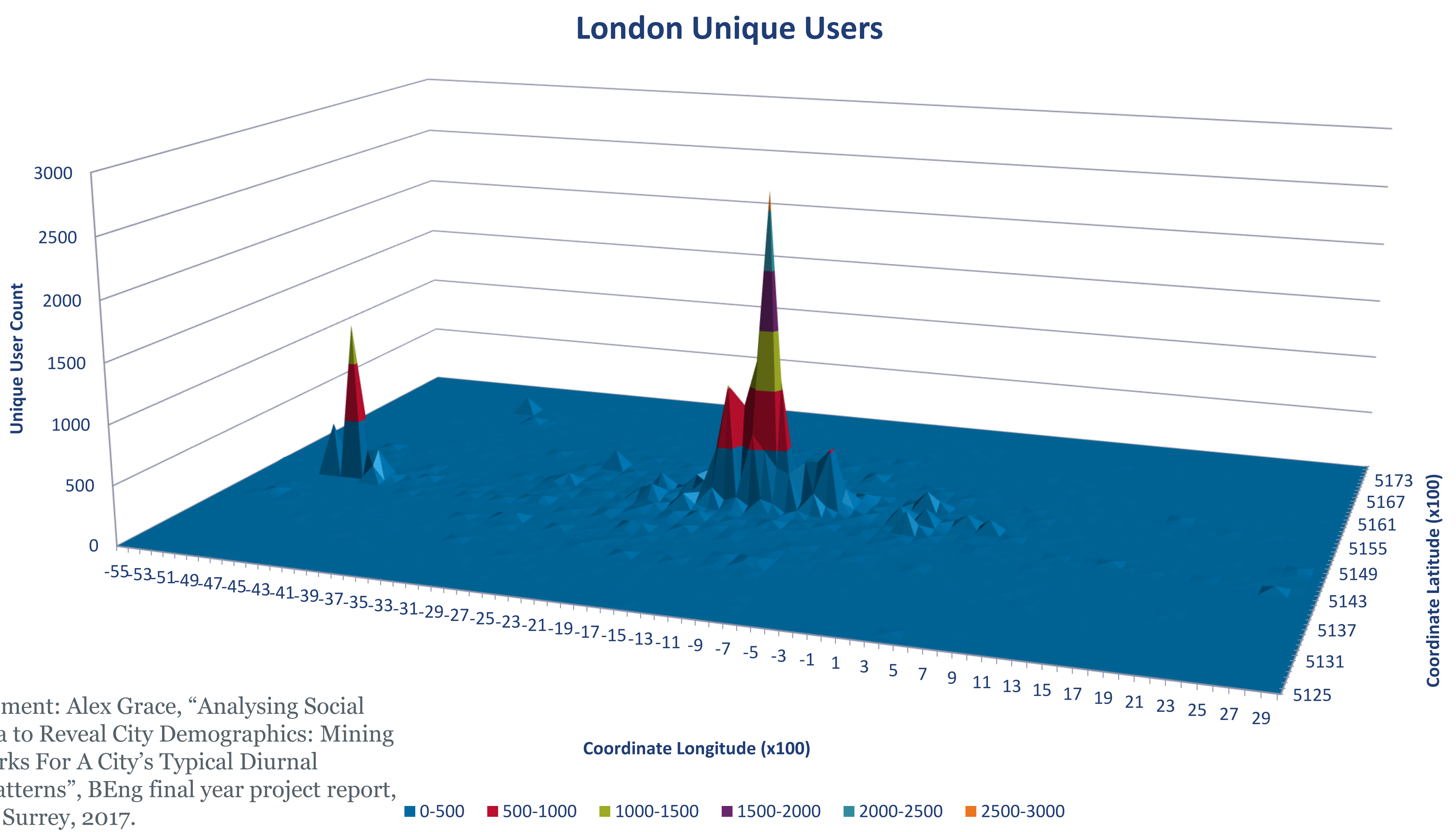
– Foursquare check-ins recorded in London and New York
  - >2 months in 2016; 39414 check-ins (London) and 98726 (NY)
  - Unique users:
    – NY – 7363
    – London – 4417
  - Venues:
    – NY – 25100
    – London – 10853



Heatmap of Foursquare venues in London

Acknowledgement: Alex Grace, “Analysing Social Network Data to Reveal City Demographics: Mining Social Networks For A City’s Typical Diurnal Movement Patterns”, BEng final year project report, University of Surrey, 2017.

CPS Data Analysis: City Rhythms



Acknowledgement: Alex Grace, “Analysing Social Network Data to Reveal City Demographics: Mining Social Networks For A City’s Typical Diurnal Movement Patterns”, BEng final year project report, University of Surrey, 2017.

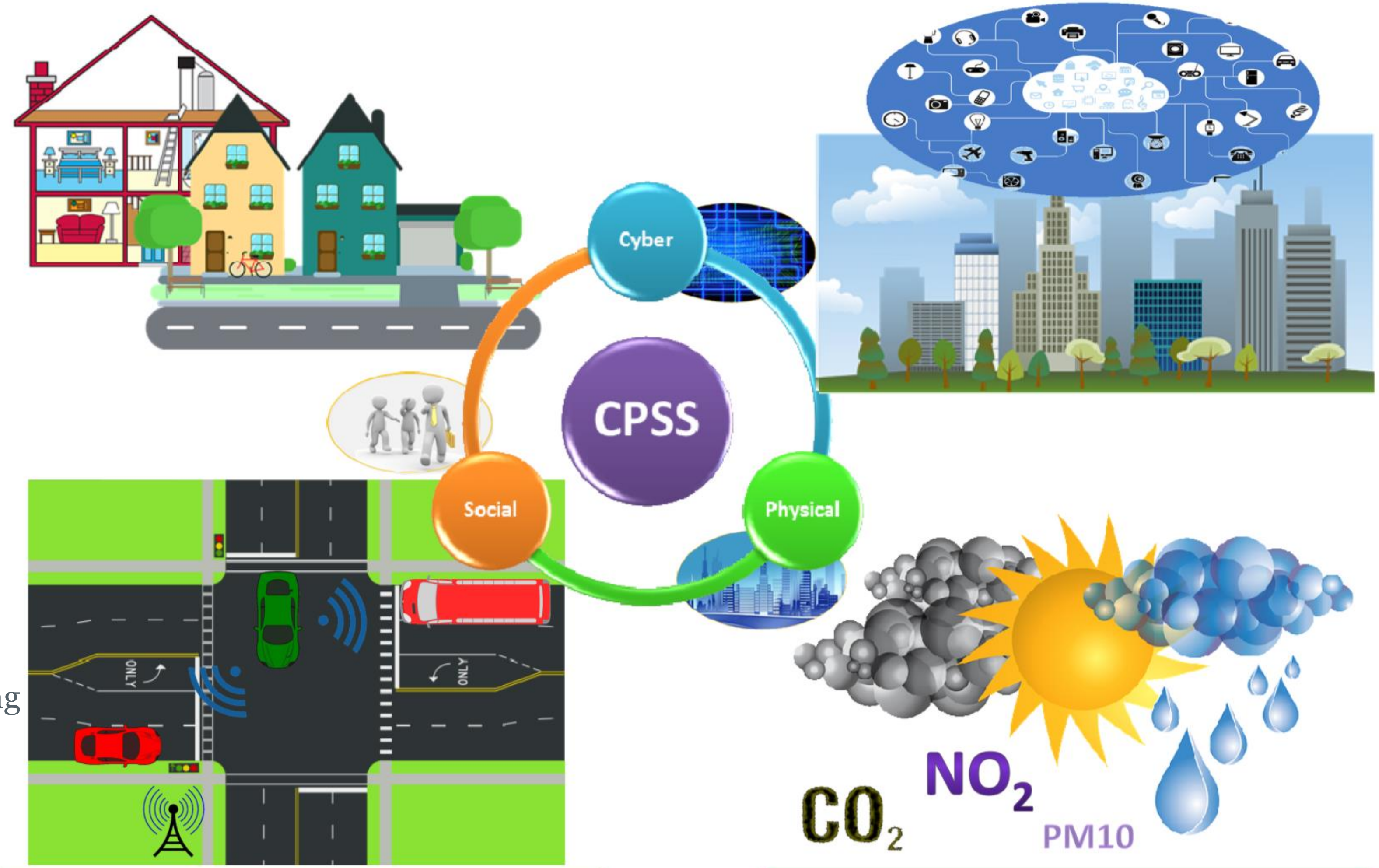
IoT Applications

Smart Home

Smart thermostat
Smart lighting
Assisted Living

Urban Intelligence

Noise heatmaps
Characterisation of urban neighbourhoods
Event/venue recommendations
Event detection
Electric vehicle charging



Intelligent Transportation System (ITS)

Smart traffic lights
Smart parking meters
Individual trip planning

Environment Monitoring

Pollution mapping and monitoring
Sentiment analysis – environment conditions

Refer to publication [5] for details.

Open Research Issues

- Cross-space data fusion:
  - Mobile crowd sensed and physical sensor data are multimodal and in different scales of measure
  - Physical sensor data in interval or ratio scale
  - Open datasets in nominal scale (qualitative classifications)
    - Need intelligent methods to convert mobile crowd sensed data into ratio scale for efficient integration with physical sensor data
- Reasoning methods: typically deductive
  - Extend to probabilistic reasoning to handle uncertain situations
  - Learning in dynamic or evolving environments
    - Detect changes in environment to trigger adaptive strategies

Selected Publications

1. De, S.; Elsaleh, T.; Barnaghi, P.; Meissner, S. An Internet of Things Platform for Real-World and Digital Objects. *Journal of Scalable Computing: Practice and Experience* **2012**, *13*, 45-57.
2. De, S.; Christophe, B.; Moessner, K. Semantic Enablers for Dynamic Digital-Physical Object Associations in a Federated Node Architecture for the Internet of Things. *Ad Hoc Networks* **2014**, *18*, 102-120.
3. Wang, W.; De, S.; Cassar, G.; Moessner, K. An experimental study on geospatial indexing for sensor service discovery. *Expert Systems with Applications* **2015**, *42*, 3528-3538.
4. De, S.; Zhou, Y.; K., M. Ontologies and context modeling for the Web of Things. In *Managing the Web of Things: Linking the Real World to the Web*, 1 ed.; Sheng M, Q.Y., Yao L, Benatallah B, Ed. Morgan Kaufmann: Burlington, Massachusetts, **2017**.
5. De, S.; Zhou, Y.; Larizgoitia Abad, I.; Moessner, K. Cyber–Physical–Social Frameworks for Urban Big Data Systems: A Survey. *Applied Sciences* **2017**, *7*, 1017.
6. Zhou, Y.; De, S.; Wang, W.; Moessner, K.; Palaniswami, M. Spatial Indexing for Data Searching in Mobile Sensing Environments. *Sensors* **2017**, *17*, 1427.
7. Komninos, A.; Stefanis, V.; Plessas, A.; Besharat, J. Capturing Urban Dynamics with Scarce Check-In Data. *Pervasive Computing, IEEE* **2013**, *12*, 20-28.

Contact

Dr. Suparna De
Senior Research Fellow
Institute for Communication Systems (ICS)
University of Surrey
Guildford, UK
Email: s.de@surrey.ac.uk
Website: www.surrey.ac.uk/ics/people/suparna_de