Horizon 2020
European Union funding
for Research & Innovation

LINDA PROJECT
LINKED DATA

ENTROPY

SubSol

UBITECH
ubiquitous solutions

Data Management, Fusion and Analytics over Heterogeneous Environmental Data

Dr. Anastasios Zafeiropoulos,
R&D Architect,
Ubitech Ltd.
azafeiropoulos@ubitech.eu

Eleni Fotopoulou,
Software Engineer,
Ubitech Ltd.
efotopoulou@ubitech.eu

MARSOL Lavrion Workshop, Athens, 16 – 18 March 2016

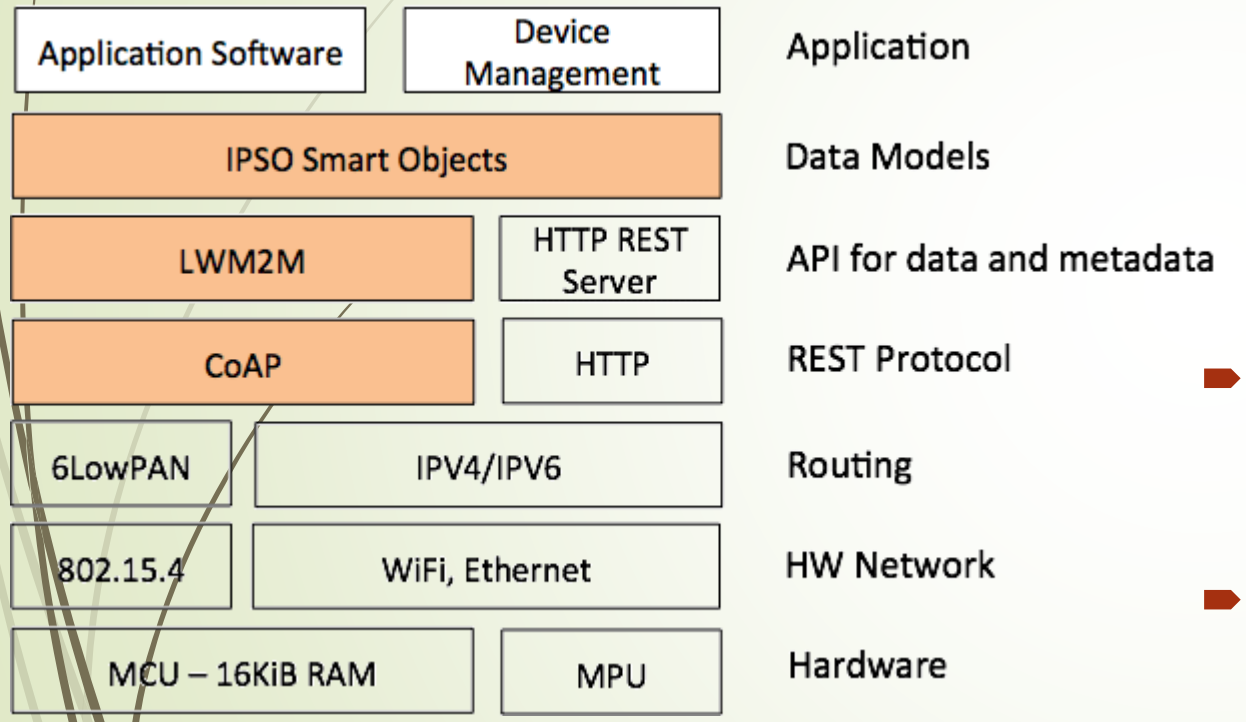
Plethora of Data

- Data coming from **Internet Of Thing (IoT)** Nodes
  - Sensor networks (e.g. environmental monitoring stations)
  - Smartphones
- Data coming from **Crowdsensing mechanisms**
  - e.g. data collected from Social Media feeds
- Data coming from available **international/national/regional databases**
- Data coming from **Satellite networks**
- Data coming from **environmental scientists/research institutes**
  - Analysis results
  - Forecasting
  - Environmental models

How to exploit the available data?

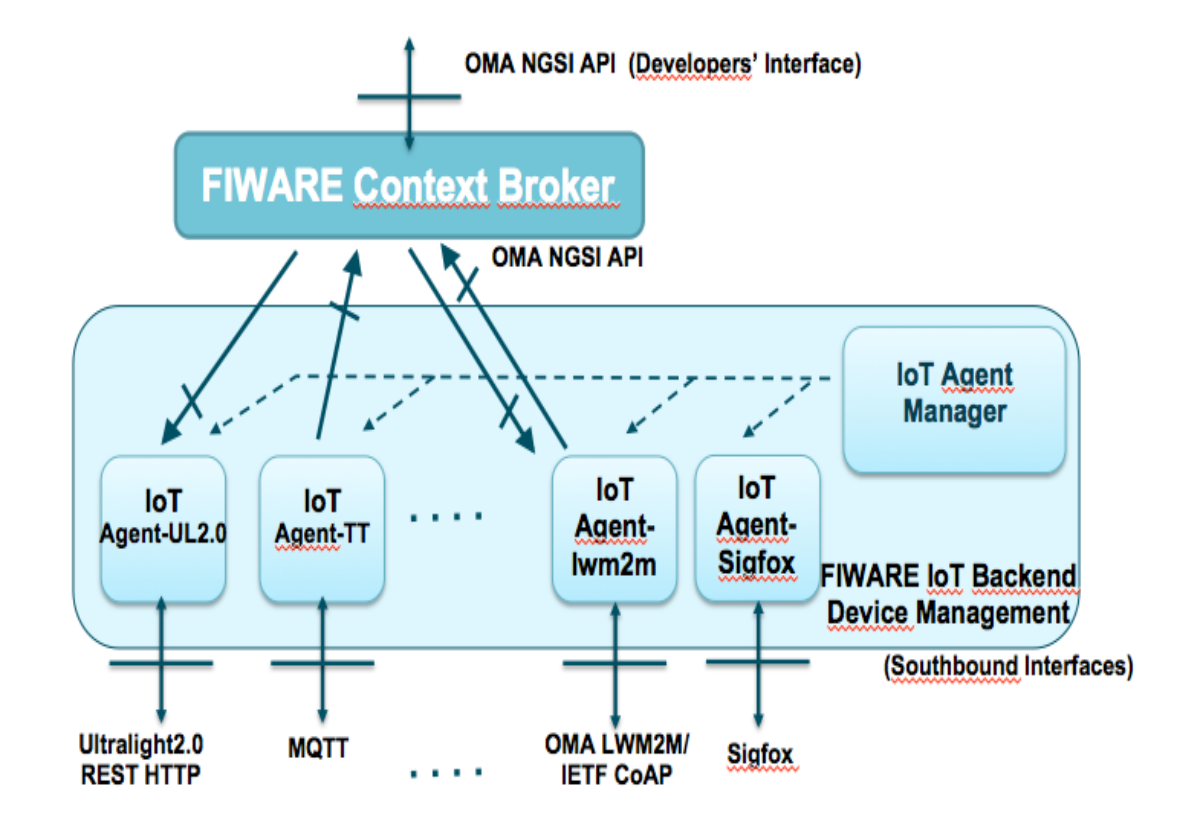
- Need for efficient and user friendly **data aggregation schemes and tools**
- Need for commonly accepted **representation models**
- Need for techniques for easily **interlinking** available data
- Need for techniques for evaluating the **data quality**
- Need for **scalable mechanisms** for data management and processing
- Need for **reasoning** over the available data
- Need for getting insights via **analytics**
- Need for **data scientists!**

Sensors Registration and Data aggregation mechanisms



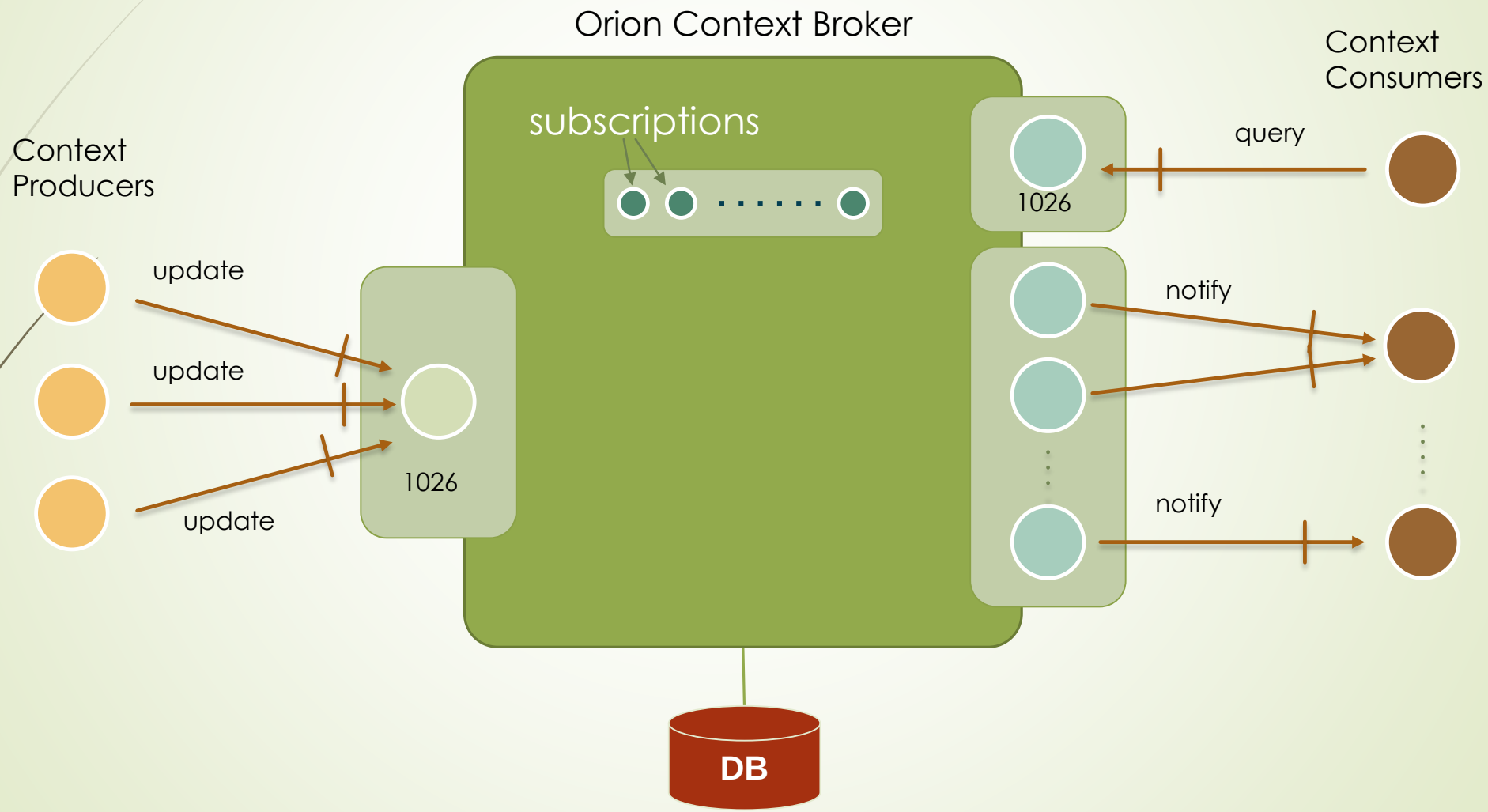
- **Lightweight machine-to-machine** (M2M) communication protocols
  - Need to support various communication standards
  - Low power communication characteristics (e.g. Zigbee, 6LowPAN)
  - LightweightM2M is to develop a fast deployable client-server specification to provide machine to machine service.
- **Device Management** (Device registration/Bootstrapping/Device configuration/Firmware update/Monitoring and Statistics)
- **Publish/Subscribe mechanisms** for data collection
  - Publish information based on set of topics
  - Consume information based on registration in specific topics

FIWARE Context Broker Overview

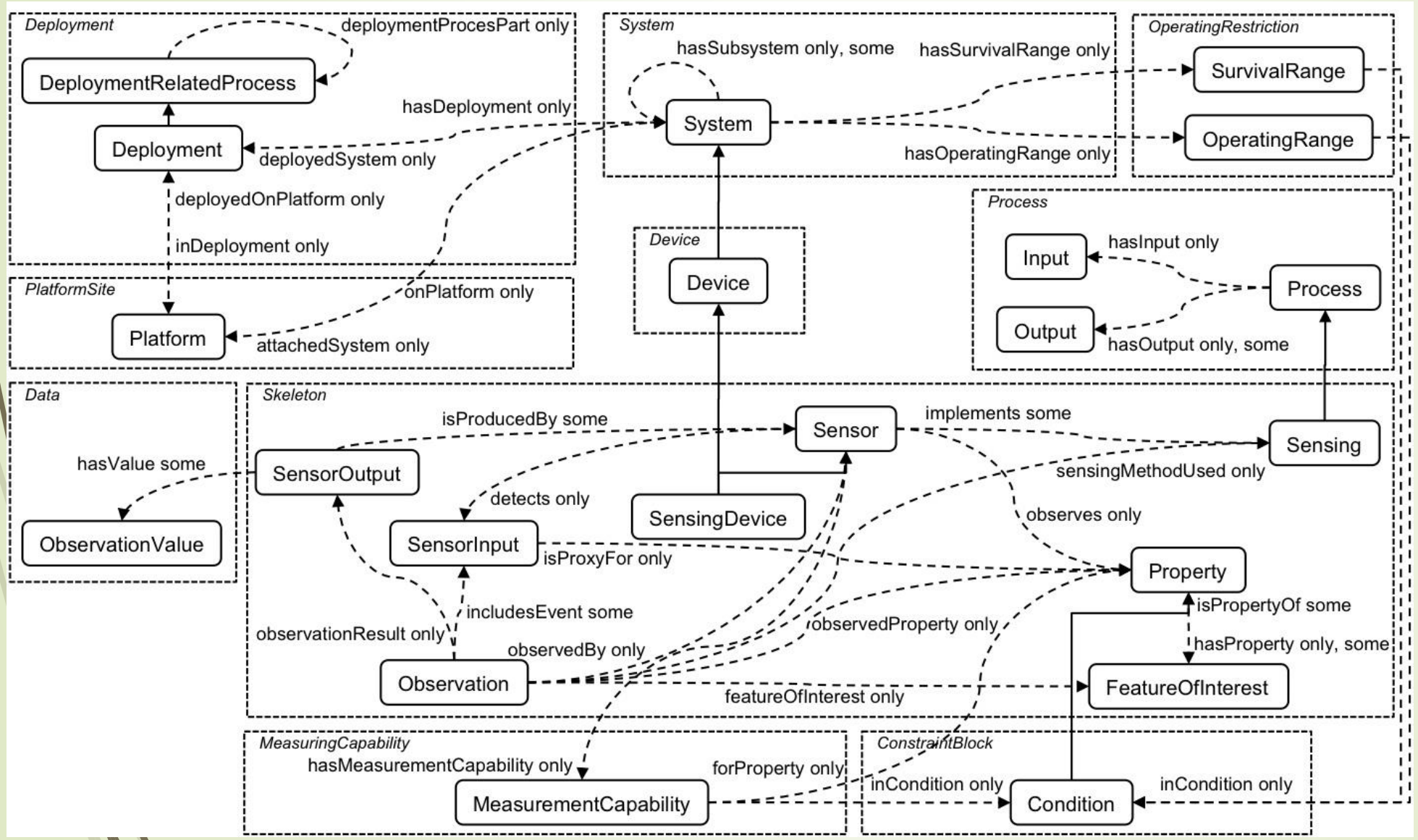


- **Register** context producer **applications**, e.g. a temperature sensor within a room
- **Update context information**, e.g. send updates of temperature
- Being **notified** when **changes** on context information **take place** (e.g. the temperature has changed) or with a given frequency (e.g. get the temperature each minute)
- **Query context information**. The Orion Context Broker stores context information updated from applications, so queries are resolved based on that information.

FIWARE Context Broker Publish/Subscribe Mechanisms



Data modeling and representation



Semantic Sensor Network Models

Applicable to many environmental specific domains

Semantic Models for Water Observations Data

- **OGC WaterML**: WaterML 2.0 is a standard information model for the representation of water observations data, with the intent of allowing the exchange of such data sets across information systems.
  - provide a common exchange format for hydrological time-series
  - build on existing standards like GML and Observations & Measurements
  - provide the option to fully store information including information regarding quality, validity/interpolation, and remarks
- **GeoSciML** version 4.0 is a data transfer standard for geological data - from basic map data up to complex relational geological databases.
- **INSPIRE Groundwater Model**: describes two basic elements:
  - the rock system (including aquifers, dependent on the geological condition) and
  - the groundwater system (including groundwater bodies), completed by hydrogeological objects (such as water wells)

OGC WaterML Specification

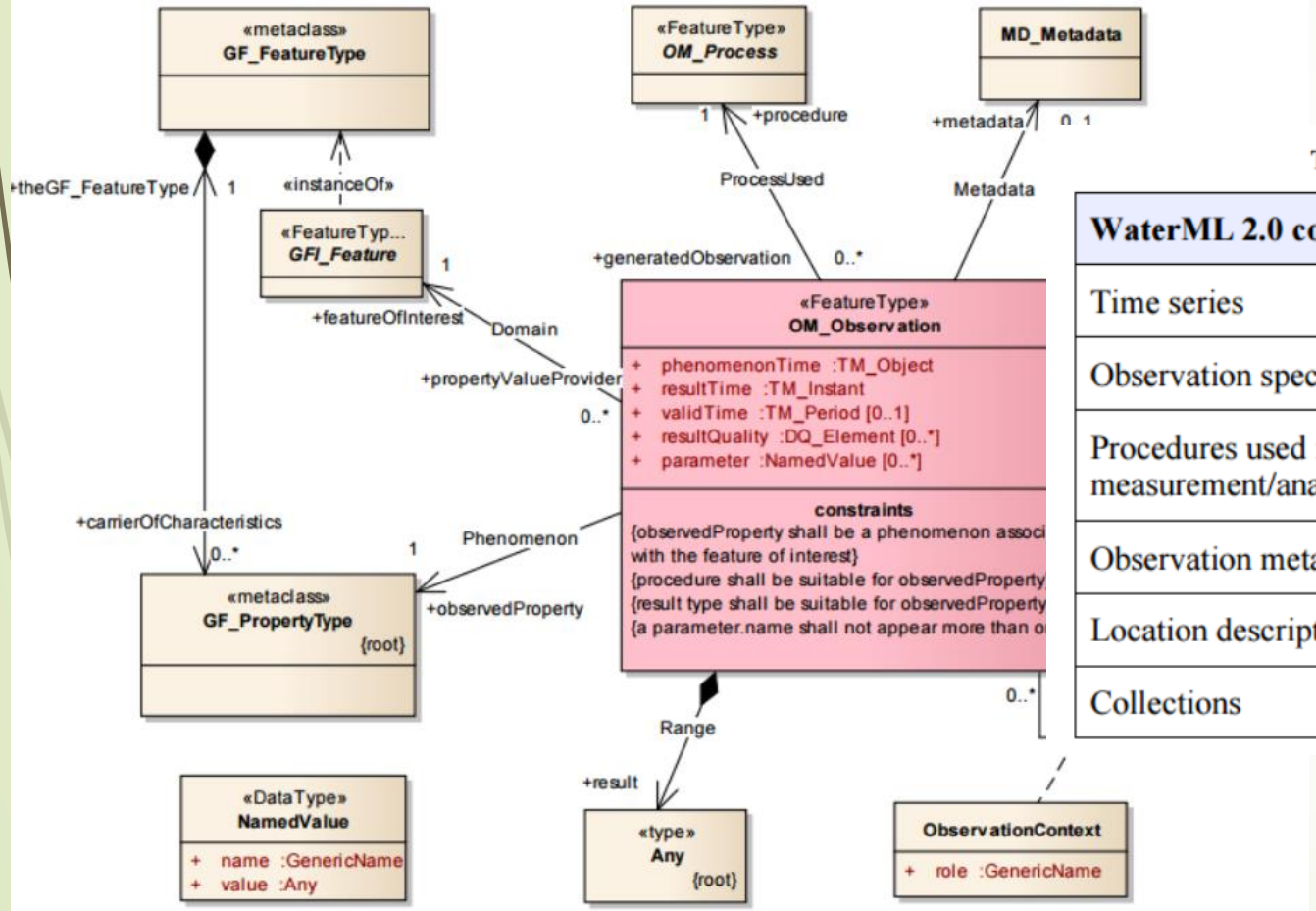
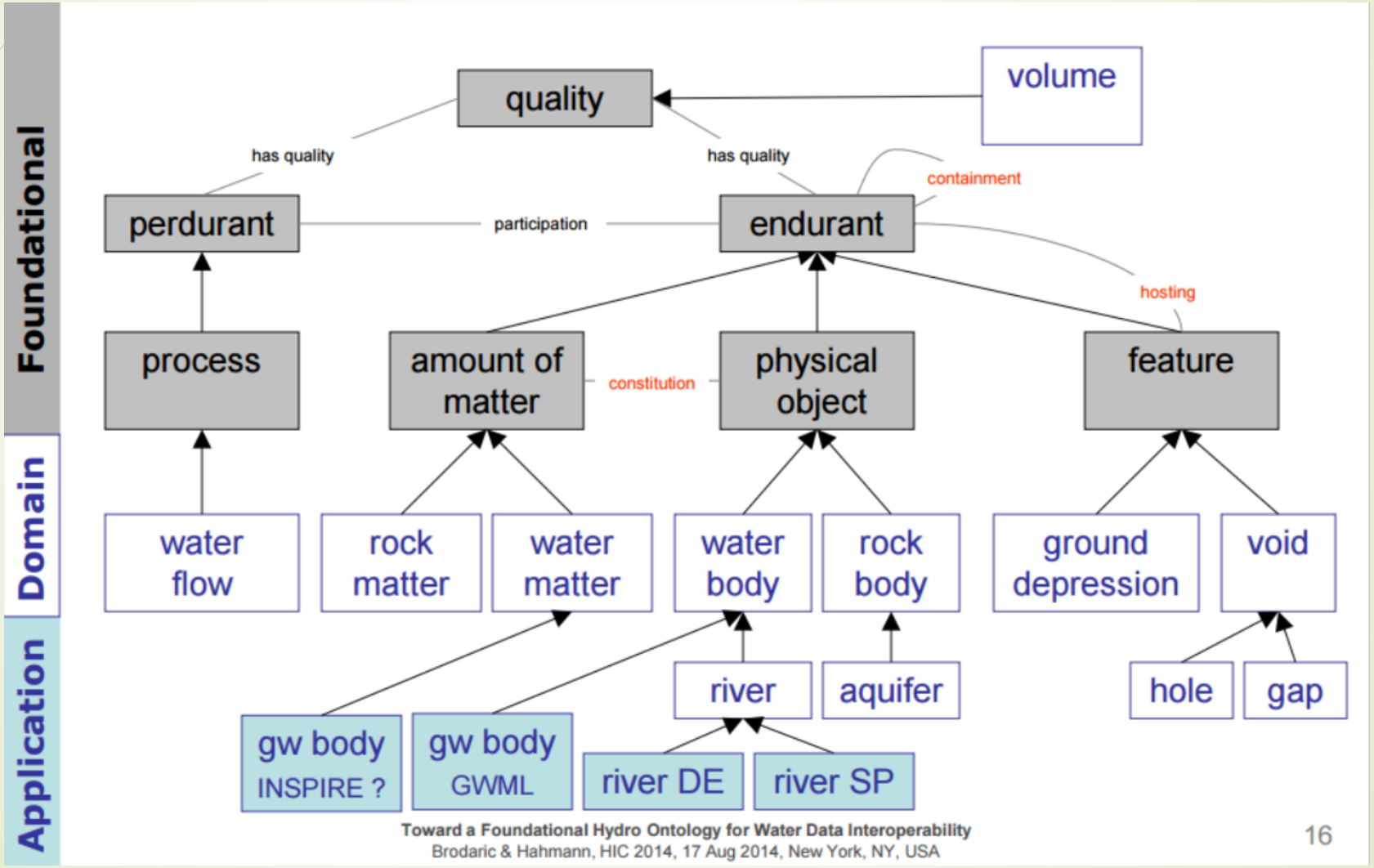


Figure 1 - Observation as defined by O&M

Table 1 - WaterML 2.0 components and equivalent concepts in O&M 2.0

WaterML 2.0 components	O & M 2.0 concepts
Time series	Result
Observation specialisations	Observation
Procedures used in measurement/analysis/processing	Procedure
Observation metadata	Observation (metadata)
Location description	Sampling features
Collections	-

Hydro Ontology for Water Data



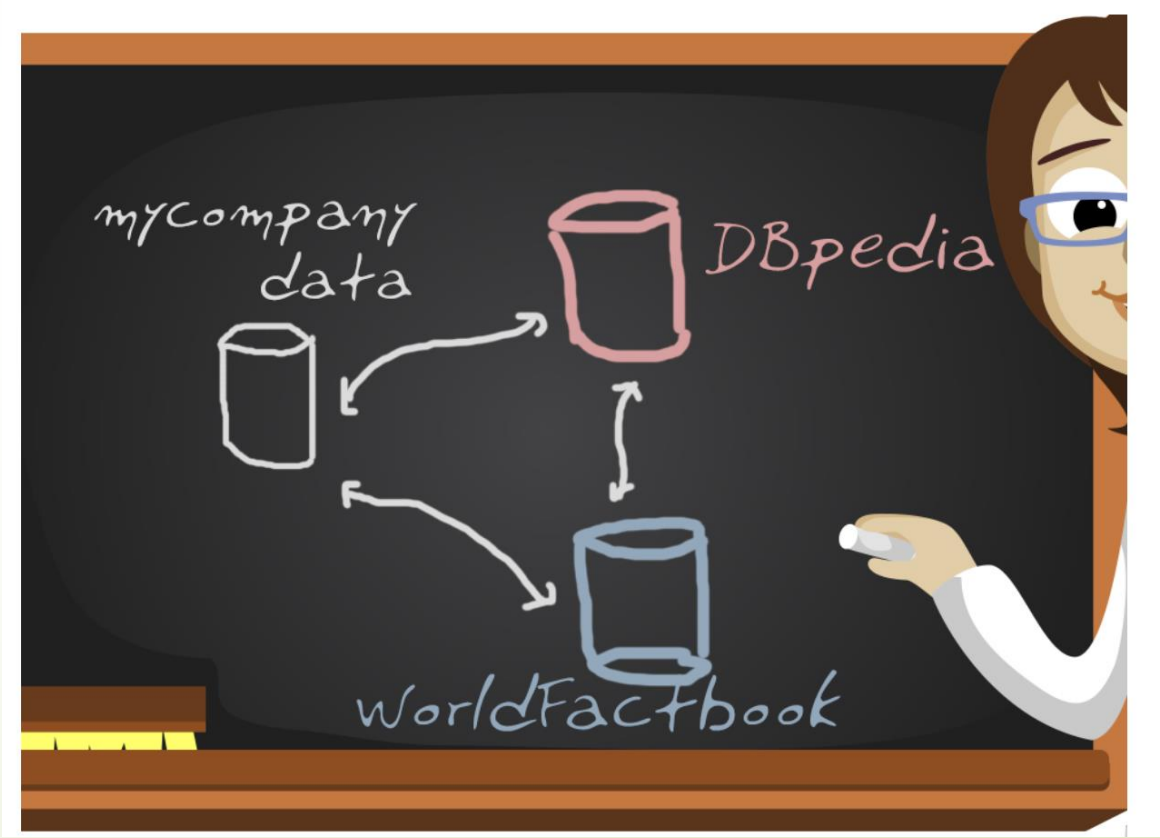
Open and Linked Data

- Open Data: publish available data usually represented in commonly used formats
- Linked Data: link data among different datasets
- LinDA FP7 project, http://linda.epu.ntua.gr/, http://linda-project.eu/

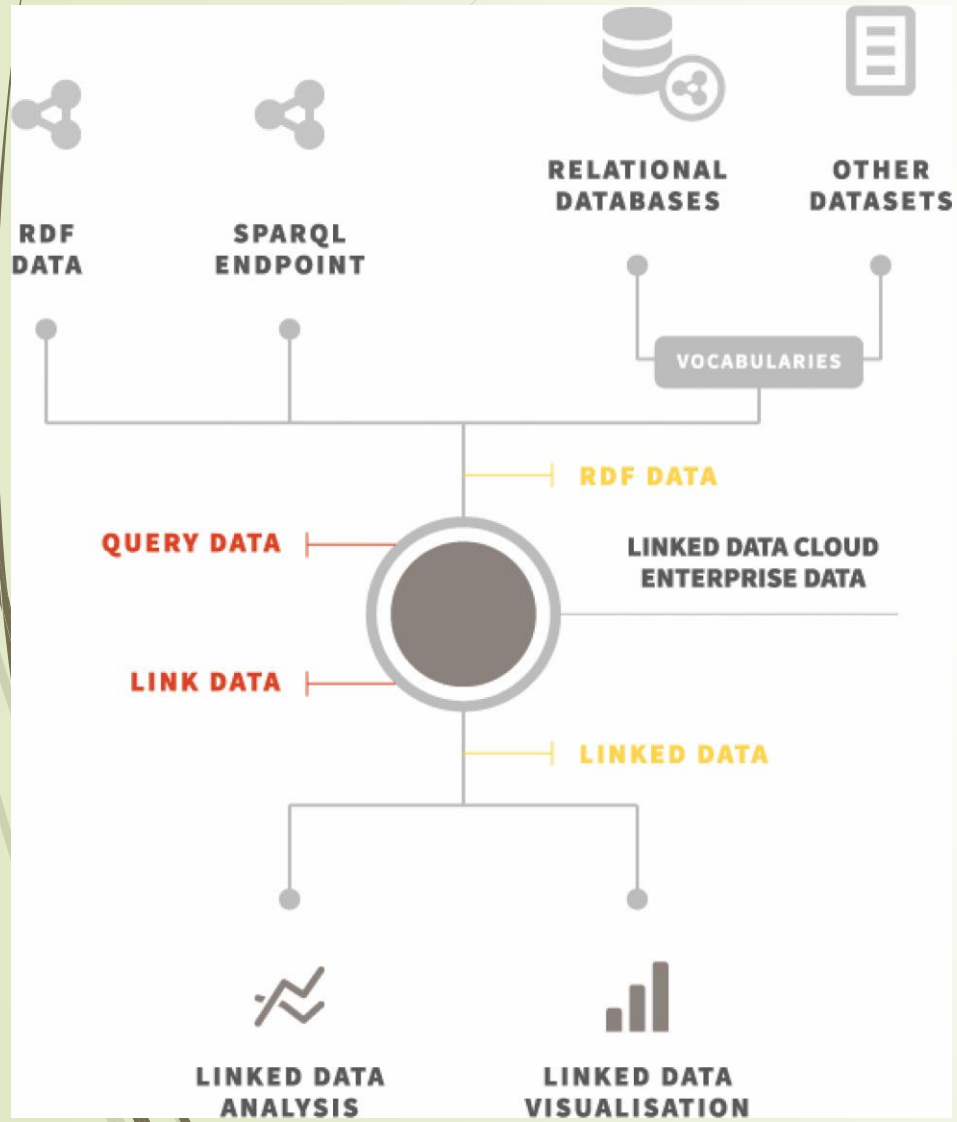
LINDA PROJECT
LINKED DATA

Power of Linked Data

- Use of URIs for data
- Schema-defying model
- Interoperability
- Interlinking of datasets



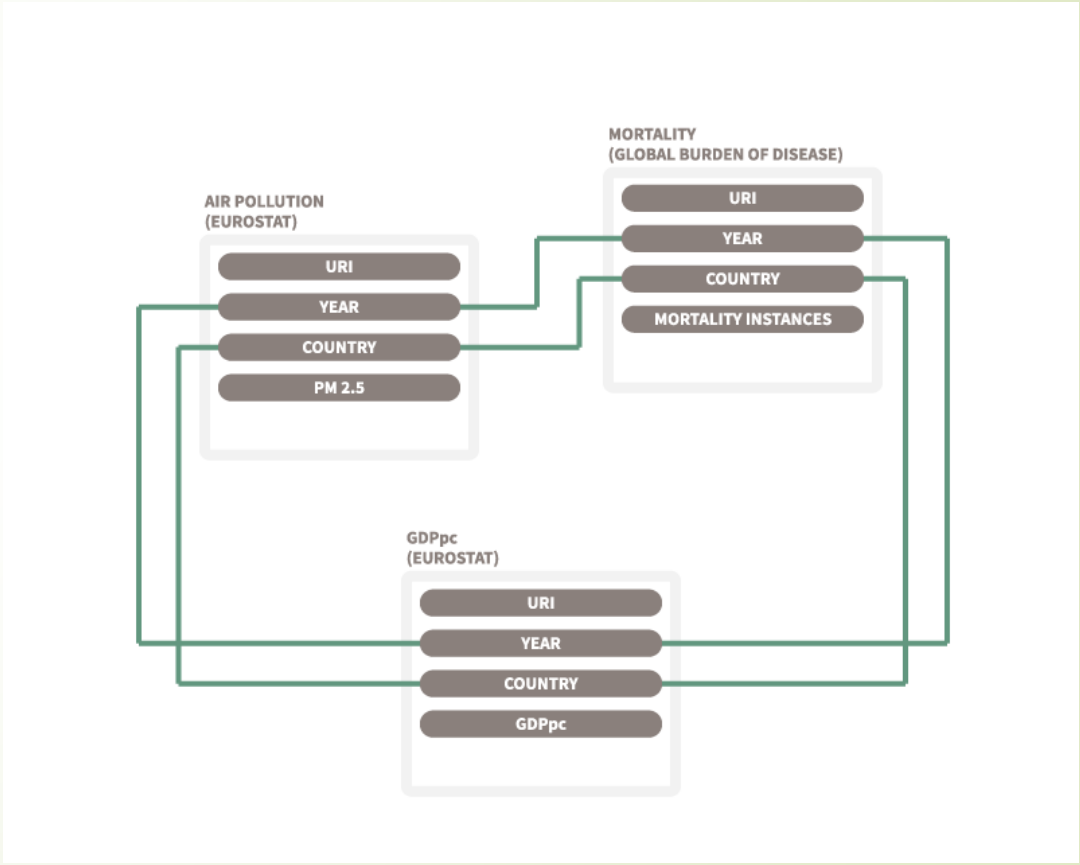
Linked Data Analytics through LinDA



Category	Supported Algorithms	Tool
Classification	J48 (decision making) M5P (piecewise linear fit to the dependent variable)	Weka
Association	Apriori (decision making)	Weka
Clustering	KMeans - Partitioning (unsupervised learning) Ward Hierarchical Agglomerative (unsupervised learning) Model Based Clustering (unsupervised learning)	R
Regression	Linear/Multiple linear regression (identify relationships and trends)	R
Forecasting	Arima (market analysis trends and seasonality patterns)	R
Geospatial	Morans I (detect spatial autocorrelation) Kriging (describe spatial autocorrelation) NCF correlogram (describe spatial autocorrelation)	R

Health Impact of Air Pollution in Urban Areas

HIGH LEVEL CONCEPTS	OUTDOOR AIR POLLUTION	HEALTH IMPACT	LIFE & ECONOMIC COST	POLICY MAKING
ANALYSIS PARAMETERS	OUTDOOR AIR POLLUTION DUE TO ROAD TRANSPORTATION, INDUSTRIAL ACTIVITY, ETC	MORTALITY, MORBIDITY	VALUE OF STATISTICAL LIFE (VSL)	LEGISLATION, DIRECTIVES, POLITICAL PARTIES POSITIONING
DATASETS IN INTERNATIONAL LEVEL	EUROSTAT, EUROPEAN ENVIRONMENTAL AGENCY	WORLD HEALTH ORGANIZATION, GLOBAL BURDEN OF DISEASE (MORTALITY & CAUSE OF DEATH)	OECD (GDP, PPP, GDP GROWTH RATES), EUROSTAT (GDP PER CAPITA), WORLDBANK	LEGISLATION, DIRECTIVES, POLITICAL PARTIES POSITIONING, TALK OF EUROPE
DATASETS IN NATIONAL & REGIONAL LEVEL	ISAT, EUROPEAN ENVIRONMENTAL AGENCY (AIRBASE), ARS TOSCANA	ISTAT, ARS TOSCANA	ISTAT, ARS TOSCANA	LINKED DATASETS OF THE ITALIAN CHAMBER OF DEPUTIES

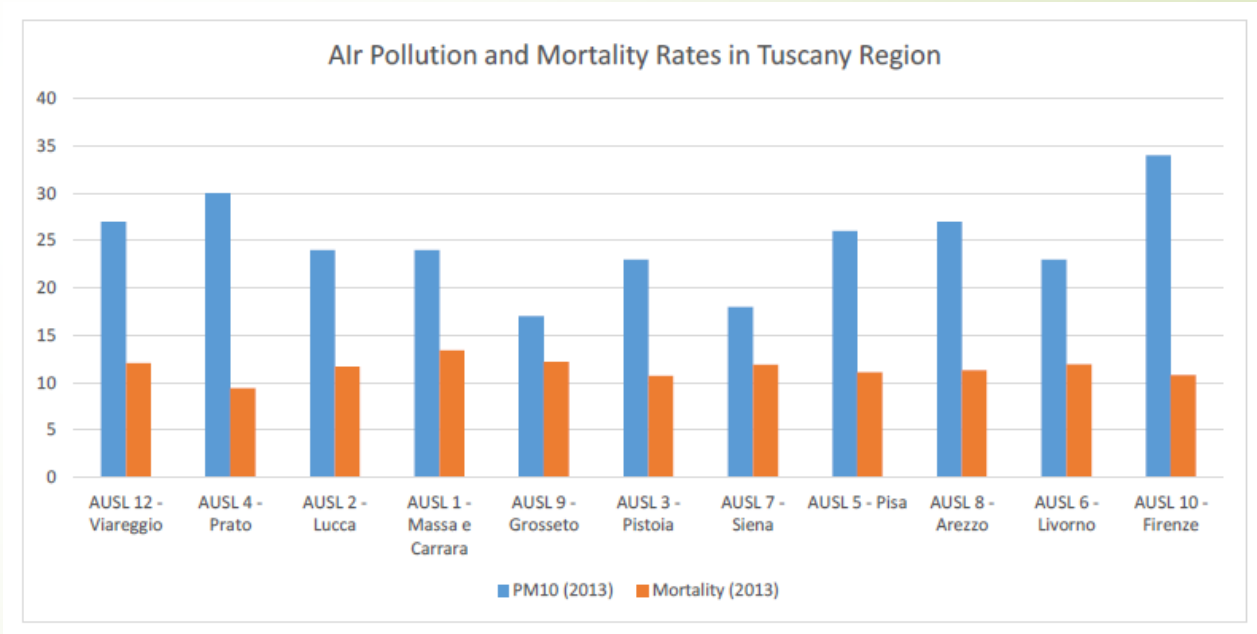


Health Impact of Air Pollution in Urban Areas

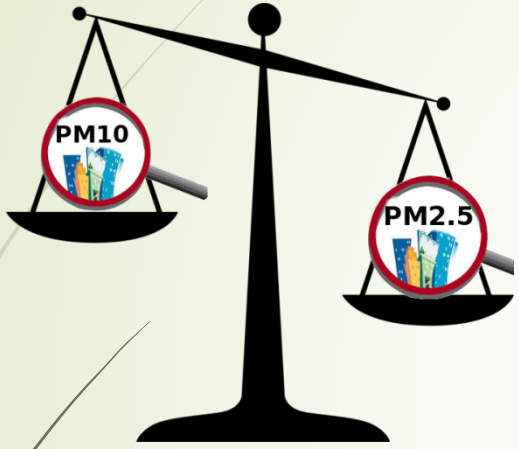
TABLE II
ANALYSIS RESULTS IN INTERNATIONAL LEVEL

Linear Model 1:
Mortality (instances) ~ PM2.5 (tons)

Disease	R-Squared	p-value	Produced Linear Regression Model Coefficients
Total Mortality	0.5635	<0.0001	21.17*PM2.5
Chr. Respiratory	0.5513	<0.0001	1.132*PM2.5
Asthma	0.4912	<0.0001	0.06889*PM2.5
Cardiovascular	0.5549	<0.0001	0.8258*PM2.5
Ischemic Heart	0.4496	<0.0001	0.4266*PM2.5
Cerebrovascular	0.5269	<0.0001	0.248*PM2.5
Diabetes	0.4503	<0.0001	0.04699*PM2.5
Trachea, Bronch. and Lung Cancer	0.5451	<0.0001	0.1176*PM2.5
Brain, nervous system Cancers	0.5876	<0.0001	0.006641*PM2.5



Health Impact of Air Pollution in Urban Areas



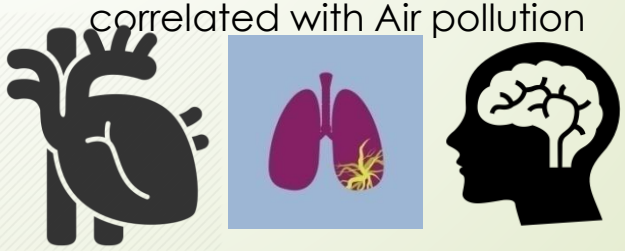
PM2.5 affect health more than PM10

Air pollution mostly affects elderly people 70+

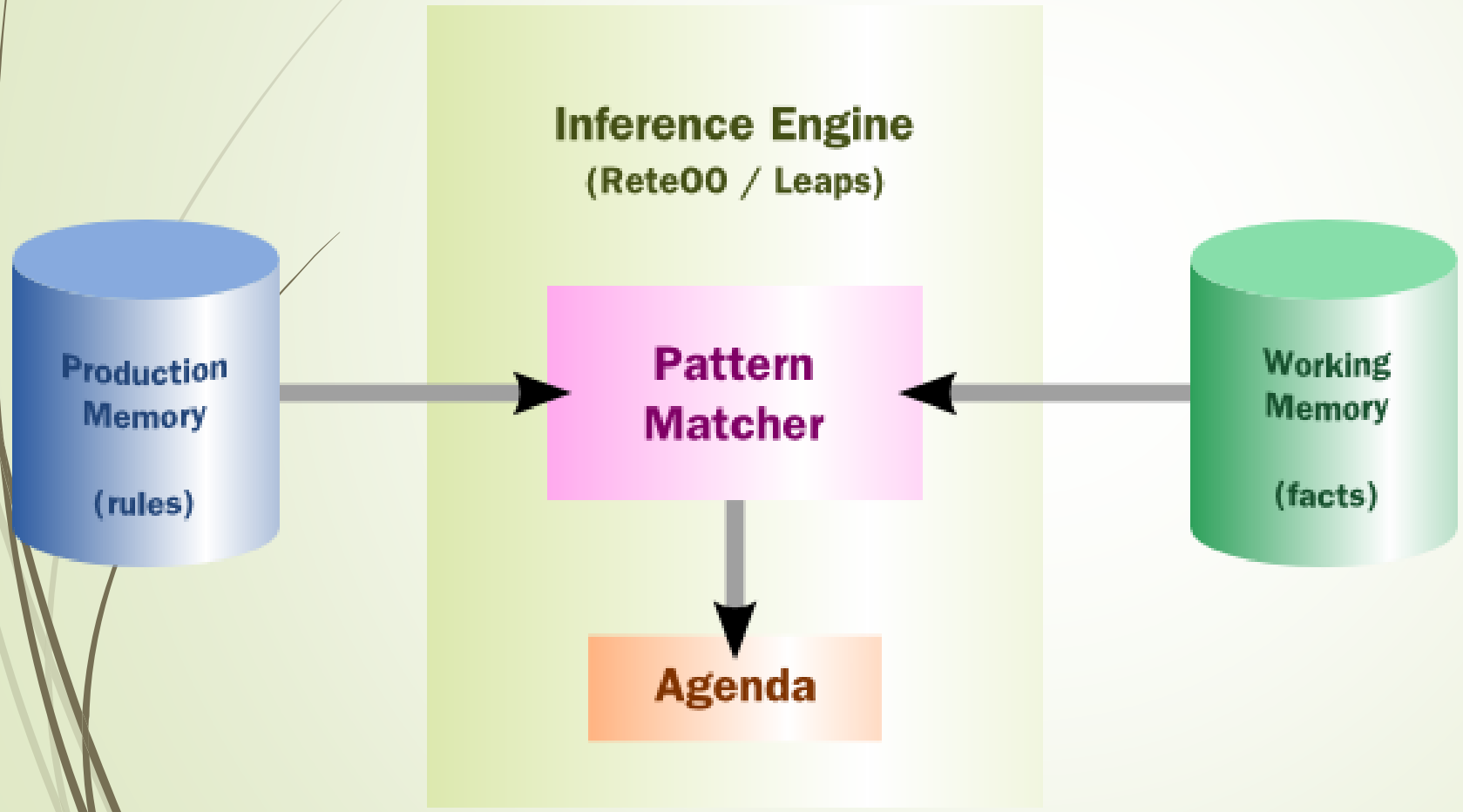
Combination of PM2.5 that come from transportation emissions and GDPpc, predicts even better the mortality instances

Air Pollution due to transportation activities affects more human health than energy production & industry sectors

Chr. Respiratory , Cardiovascular, Cerebrovascular , Brain, nervous system Cancers are the diseases that are mostly correlated with Air pollution



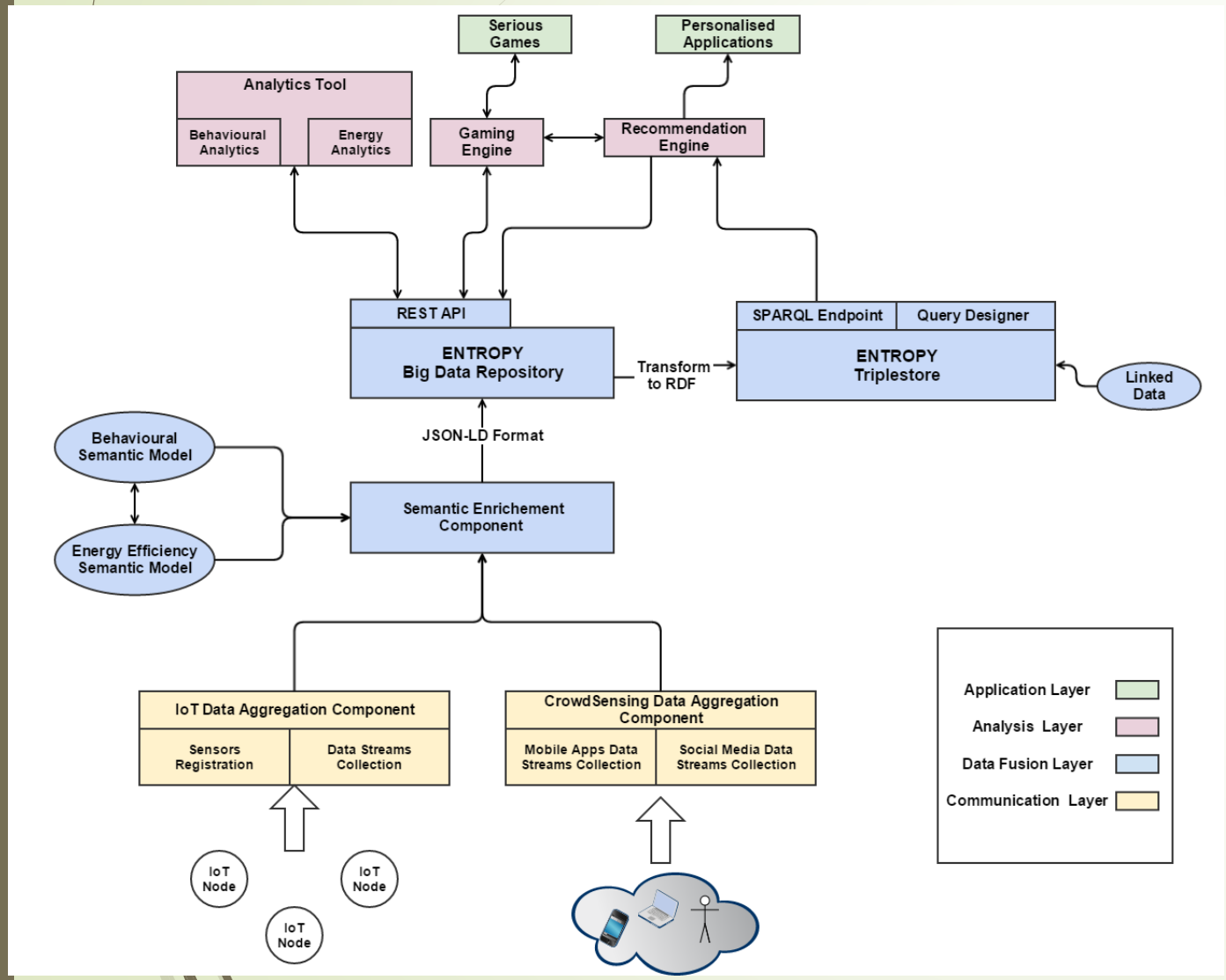
Reasoning over Semantic Modeled Data – Rule Engine and Production Rule System

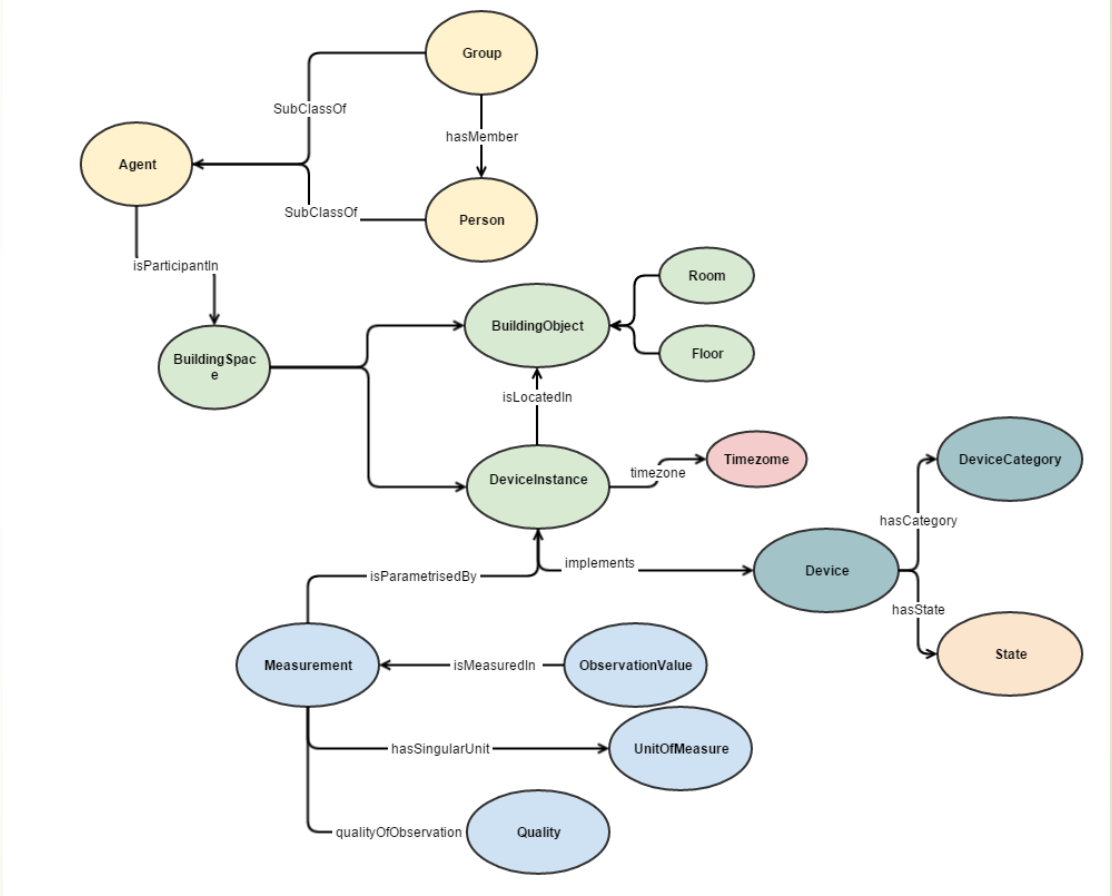


Municipal water management solution

- identification of faults, alarms;
- manage water distribution;
- adapt pricing policies (smart grid oriented concepts)

Improving Energy Efficiency in Smart Buildings through Behavioural Change



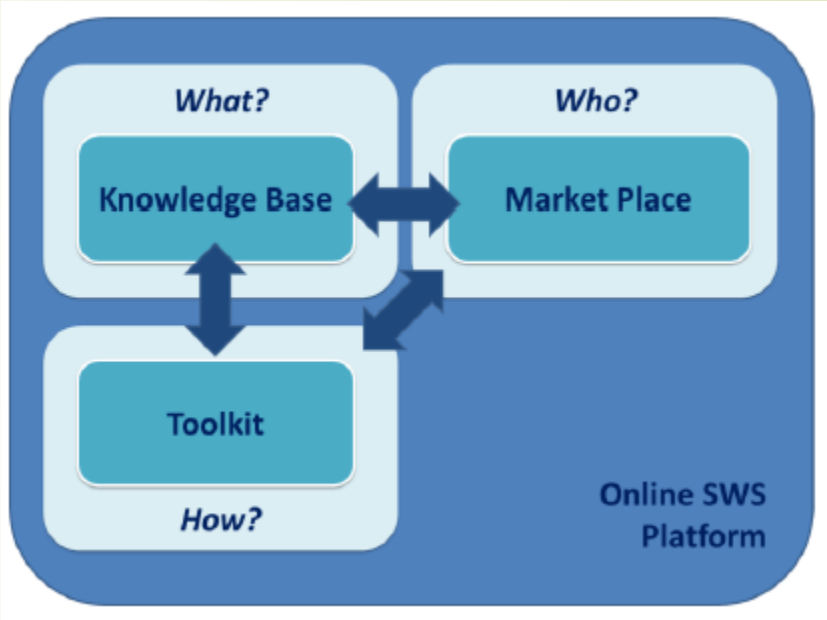


ENTROPY

ENTROPY H2020 Project,
http://entropy-project.eu/

Water Data Management in SUBSOL

- Data collection per considered installation site
  - Data from IoT nodes
  - Data from meteo stations
  - Data from crowdsensing mechanisms
  - SUBSOL Knowledge Base
- Data Visualisations and Analytics Toolkit
  - dashboard available per installation site
  - support of set of data mining algorithms
  - exploitation of linked data technologies where considered beneficial
- Water Management Solutions Market Place
  - comparisons among installations
  - dissemination of best practices
  - products/services market place



Thank your for your attention!

Contact: azafeiropoulos@ubitech.eu