International Journal on Semantic Web and Information Systems (IJSWIS)
Volume 13 • Issue 3 • Article 1 • 2017

SEMDPA: A Semantic Web Crossroad Architecture for WSNs in the Internet of Things

Eliot Bytyçi
University of Prishtina

Besmir Sejdiu*
University of Prishtina

Arten Avdiu
South East European University

Lule Ahmedi
University of Prishtina

ABSTRACT

The Internet of Things (IoT) vision is to connect uniquely identifiable devices that surround us to the Internet, which is best described through ontologies. Thereby, new emerging technologies such as wireless sensor networks (WSN) are recognized as an essential enabling component of the IoT today. Hence, given the increasing interest to provide linked sensor data through the Web either following the Semantic Web Enablement (SWE) standard or the Linked Data approach, there is a need to also explore those data for potential hidden knowledge through data mining techniques utilized by a domain ontology. Following that rationale, a new lightweight IoT architecture SEMDPA has been developed. It supports linking sensors and other devices, as well as people via a single web by mean of a device-person-activity (DPA) crossroad ontology. The architecture is validated by mean of three rich-in-semantic services: contextual data mining over WSN, semantic WSN web enablement, and Linked WSN data. SEMDPA could be easily extensible to capture semantics of input sensor data from other domains as well.

Keywords: Internet of Things, Ontologies, Data Mining, Sensor Web Enablement, Linked Open Data

INTRODUCTION

Internet of Things (IoT) paradigm has been around since almost two decades but its meaning has undergone significant changes. Initially, the term was though as a way to link supply chain with radio frequency identification (RFID) (Ashton, 2009). Nowadays, IoT vision is to connect to the Internet not only computers, tablets or smartphones but also other physical objects and devices surrounding us, such as sensors, actuators, etc. – which, through unique identifiers are able to interact with each other to reach common goals in everyday life like in environmental monitoring, e-health, domotics, or in automation and industrial manufacturing just to mention few (Atzori, 2010), (Giusto, 2010).

That would set up a *triangle device – person - activity* with relations drawn between people and devices. It should be further explored in order to infer important knowledge like are activities of people but also their devices. All of that needs meaning interpretation, which can be obtained by semantics. Therefore, using *ontologies* to describe the conceptualization of this certain domain is necessary. Some researchers argue that sensor networks are the most essential components of the IoT, with most of the sensors today deployed as wireless (Perera, 2014). According to a recent

BBC report, the global market for sensors is expected to grow fast to almost double by 2021, with wireless sensor devices[1] nearly triple its current market by 2021. Ericsson[2] on the other side predicts that IoT sensors and devices are expected to exceed mobile phones as the largest category of connected devices in 2018. Hence, new emerging technologies such as *WSN*s (wireless sensor networks) are an imperative in this domain conceptualization. That may implicate adding new ontological constructs and constraints on top of the existing ontologies.

There is a standard, the *Sensor Web Enablement* (SWE) (Lefort, 2011), (Bröring, 2012), conceived by the Open Geospatial Consortium[3] (OCG) that supports publication on the Web of (potentially heterogeneous) WSN related data following *a single standard schema*. WSN data may thus get accessed with ease via a single web, the so-called Web Enabled WSN (Rouached 2012) (Udayakumar, 2012), leading provisionally to lower cost and better quality communication between sensors.

On the other side, there is an ever growing vast amount of data from a diversity of domains being published as *Linked Data* (Bizer, 2009), (Lee, 2006), namely *open semantically described and interlinked* data that are made available for access through the Web. It may thus be easier and time effective to build applications using Linked Data. Moreover, applications *may gain from new knowledge potentially derived from semantic descriptions including interlinks*. As more open rich-in-semantics linked sensor data are published on the Web, best practices are evolving too, such as the proposed six step model to create and publish linked data (Hyland, 2011, Cygankia, 2014) and as well the best practices for publishing Linked Data published by World Wide Web Consortium (W3C)[4].

Finally, given the increasing interest to provide linked sensor data through the Web following either the SWE standard or the Linked Data approach, there is obviously also need to further explore such data in order to check for potentially new knowledge. *Data mining* techniques may aid in *exploring massive datasets*, especially if *utilized by ontologies* imposing certain modeling on data and their semantic annotations.

The work presented here is part of a major project InWaterSense[5] (Ahmedi, 2013), which consists of a wireless sensor network (WSN) deployed in a river in Kosovo for monitoring its water quality. WSN has two main components: static component – deployed in a specific location at the river, and the mobile component – used to measure in different locations throughout the river. Both components contain several sensors, which measure different quality parameters such as: pH, temperature, dissolved oxygen, etc. As an umbrella of the project, an environmental monitoring portal was introduced. The portal supports modeling and management of both, the observational stream data on water quality coming from wireless sensors – dynamic data, as well

[1] http://www.bccresearch.com/market-research/instrumentation-and-sensors/wireless-sensors-technologies-report-ias019c.html
[2] https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf
[3] www.opengeospatial.org/
[4] https://www.w3.org/TR/ld-bp/
[5] http://inwatersense.uni-pr.edu/

as of the data describing the WSN itself, its devices and the corresponding site allocation data – static data (Ahmedi, 2015).

Following the trends and rationale discussed above, as well as the experience gained while working with the real-life WSN system within the INWATERSENSE project, a new lightweight IoT *architecture* SEMDPA has been developed and is presented in this paper. It supports linking sensors and other devices, as well as people at the input via a single web by mean of a *d*evice-*person-activity* ontology referred to as DPA. At the output, multiple services are readily supported, here exemplified through three typical services, contextual data mining over WSN, semantic WSN web enablement, and linked WSN data. The main contribution behind the proposed SEMDPA architecture with the ontology in the middle acting as a Semantic Web (Lee, 2001) crossroad may thus be summarized into:

- The *ease of service provision* via a *single schema* Web in conformity to the existing *SWE* standards which base on sensor data but also on the participating people as actors with certain activities related to sensors at the input.
- The *richness* of service provision due to *semantics* encapsulated *by mean of ontology*.
- The *support for multiple diverse services at the output*, from *data mining* to *web services* to *linked data* and even more services via a common Web.

The remainder of this article is organized as follows. In the next section, related work is discussed, followed by a section on the proposed IoT architecture SEMDPA. After that, the DPA ontology describing the concepts presented in the architecture, i.e. devices, people, as well as activities relating them is introduced in detail. Then, the platform for the IoT with its three sample rich-in-semantic services is provided. Finally, in last part, conclusions and future work are discussed.

RELATED WORK

In support of working with sensor data, *semantic* web has already been utilized to enable rich modeling and querying or even reasoning over sensor data annotated with meta-descriptions in form of ontologies: In (Calbimonte, 2011), the SSN[6] and SWEET[7] *ontologies* are used to model sensor data and to allow a federated query system among them. In (Phuoc, 2011), a Linked Stream Middleware (LSM) provides wrappers for real time data collecting and publishing, a web interface to publish data and a SPARQL[8] endpoint for querying sensor data. As part of the INWATERSENSE project, in (Ahmedi, 2013), the INWS[9] ontology which builds on top of the SSN[10] ontology models WSNs for water quality monitoring, whereas in (Jajaga, 2017a, Jajaga 2016) and (Jajaga, 2017b), a reasoning framework uses a Jess production rule system or a Semantic Web rule language C-SWRL respectively over the INWS's stream sensor data. In (Keßler, 2010), linking sensor data using *Linked Data* principles is seen as promising approach in order to make data available to users

[6] http://purl.oclc.org/NET/ssnx/ssn
[7] http://sweet.jpl.nasa.gov/2.3/sweetAll.owl
[8] http://www.w3.org/TR/sparql11-query/
[9] http://inwatersense.uni-pr.edu/ontologies/inws-core.owl
[10] http://purl.oclc.org/NET/ssnx/ssn

that are not in line with SWE standards. Even though it makes querying more difficult, by enabling annotations with timestamp and location, still it makes explicit what meta-data describes.

Apart from semantic web alone, combination of semantics and data mining, or semantics and SWE, or SWE and data mining have all been explored as well, as is recalled next.

In (Aggarwal, 2013), bringing together *semantic* web and *data mining* in the context of IoT with a focus on sensors as interconnected devices is presented. Sensors produce vast amount of data, which need to be linked first, and then described in a standardized way by reusing existing infrastructure and in the end analyzing the data. By this, authors conclude that practical data mining applications can be build, since real world sensors ontologies exist, query mechanisms as well and the availability of linked sensor data.

In (Sheth, 2008), semantic sensor web (SSW) is described as a synthesis of sensor data and *semantic* metadata. SSW represents an approach by Open Geospatial Consortium (OGC) and Semantic Web Activity of the World Wide Web Consortium (W3C) to provide meaning for sensor data. Core suit of services developed and maintained by OGC under *SWE* framework comprises of Observations and Measurements (O&M), Sensor Model Language (SML), Sensor Planning Service (SPS) and Sensor Observation Service (SOS) amongst other. In (Henson, 2009), a construction of a Semantic Sensor Observation Service (SemSOS) based on the SWE standards is discussed. They have modelled the domain of sensors and sensor observations in a suite of ontologies, adding semantic annotations to the sensor data, using the ontology models to reason over sensor observations, and extended an open source SOS implementation with their semantic knowledge base. They extend the open source implementation of SOS from 52North[11]. 52North's SOS is an open source software implementation of the Open Geospatial Consortium's Sensor Observation Service standard[12], which is designed to be highly modular, and adaptable to arbitrary suitable sensor data sources, transport protocols, etc.

In (Lee, 2015), an extension of the *SWE* framework in order to support standardized access to sensor data is described. The system also introduces a web-based data visualization and statistical analysis service for data stored in the Sensor Observation Service (SOS) by integrating open source technologies such as WEKA[13] API for *data mining* tasks. Furthermore, as future work they list the extension of SOS server with a semantics, since the lack of semantically rich mechanism is seen as a significant issue, which makes it hard to explore related concepts, subgroups of sensor types, or other dependencies between the sensors and the data they collect. Authors believe that by integrating SOS with semantics will enable for querying high-level knowledge of the environment as well as the raw sensor data. Moreover, that would also facilitate knowledge sharing and exchange, and automated processing of web resources. In (Conover, 2008), Sensor Management for Applied Research Technologies (SMART), a project developed to present the capabilities of OGC SWE for observation and forecasting is presented. A major component of the system is

[11] 52north.org
[12] http://52north.org/swe
[13] www.cs.waikato.ac.nz/ml/weka

Phenomena Extraction Algorithm (PEA), a data mining algorithm for detection of anomalies amongst other.

To the best of our knowledge, there is yet no combination of all three abovementioned ground concepts, namely of semantic web (including linked data), SWE, and data mining explored at one place for harnessing sensor data or any other data produced by devices people have interest to relate to in the context of the IoT, which is what characterizes the approach presented in this paper.

PROPOSED SEMANTIC DEVICE-PERSON-ACTIVITY (SEMDPA) ARCHITECTURE

Semantic Device-Person-Activity architecture named SEMDPA is presented in Figure 1. In the core of the architecture lies the ontology, which describes the main concepts involved: devices, people and their activities. The ontology may get fed by a WSN deployed, like for example in a river. Every person and device involved, has a specific Uniform Resource Identifier (URI) or RFID tag, which can link to other relevant information about the person or device.

In a perfect scenario, measurements from sensors, after collected, are to be validated and presented to the public. The validation should be performed by the corresponding agency responsible for environmental monitoring, Of course, in cases where there are non-validated data, the system should not transmit those to the public. The Semantic Web Enablement (SWE) plays an important role in not letting that happen. It links through ontology to the sensor data, and then further, through the *SOS web services*, describes the way sensor data are presented to the Web. Furthermore, our architecture extends on *data mining* of sensor data, through the addition of *context ontologies* for the given domain, in this case the water domain and its quality sensing. Besides, *Linked Data* further provides an additional value to the system by making sensor data annotated with semantics available to all interested parties, regardless of their OGC SWE standards conformity.

Even though the architecture is showcased for the water domain, it is *generic and adaptable to any other domain* like e-health, domotics, or automation and industrial manufacturing just to mention few, with devices and people interacting towards gathering and processing data for the sake of useful rich-in-semantics services at the output. Taking healthcare systems as an example of another domain with sensors acquiring useful data about patients, therapies, pharmaceutics, and doctors, including also people interacting with the system, the proposed architecture may at the output support provision of services like patient actual health, most expertise of doctors, certain pharmaceutics in doubt for patients' worst health, and similar.

Hence, the DPA ontology presented is aimed to serve as a Semantic Web crossroad in the middle of the architecture which supports modeling and semantically annotating and interlinking any input data of three DPA (device-person-activity) parties and in turn enables distinct rich-in-semantics services at the output. Although the architecture illustrated assumes WSN in the place of devices, it is *generic and adaptable for any arbitrary device* at the input in the context of the IoT infrastructure.

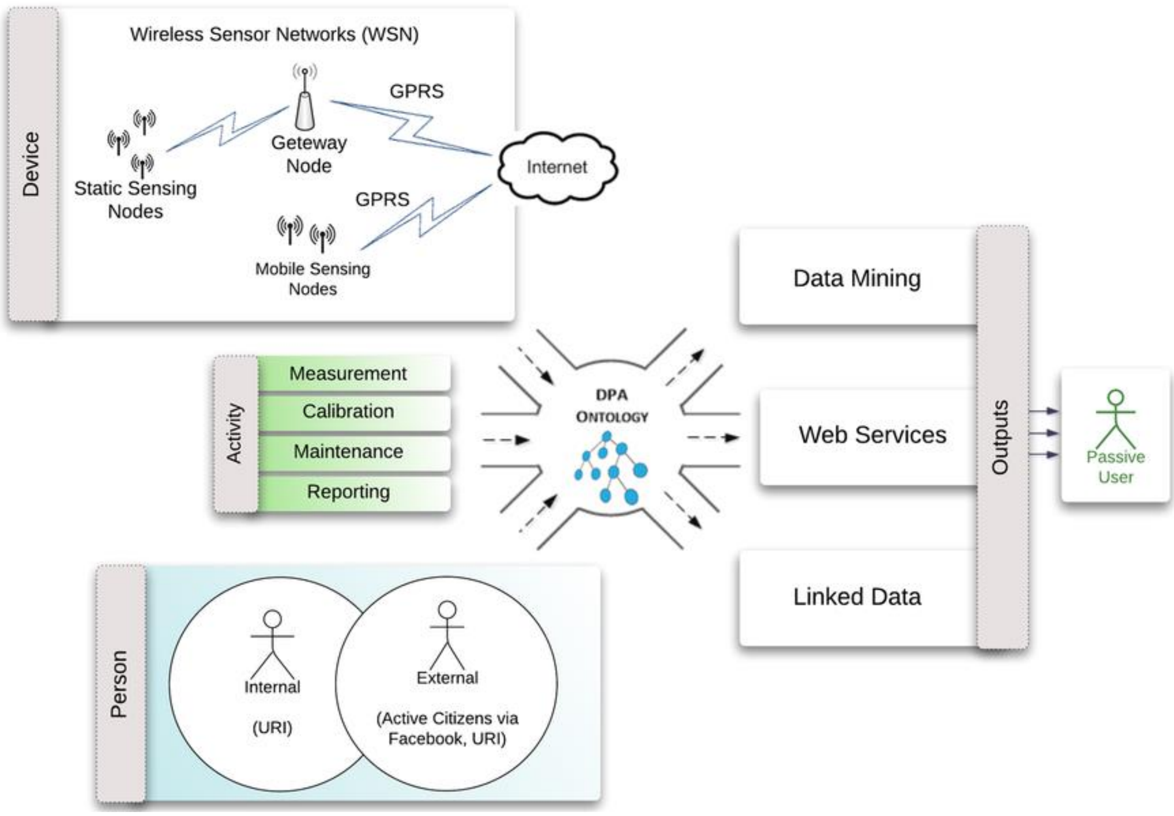

Figure 1. SEMDPA Architecture

THE DEVICE-PERSON-ACTIVITY (DPA) ONTOLOGY

Concepts belonging to the SEMDPA architecture are formally specified through the DPA ontology, presented in Figure 2, which describes relations between Device, Person and Activity, and as well time and place of the occurrence of such encounter. The ontology presents an extension of the INWATERSENSE project ontology – INWS ontology (hence of the SSN ontology as well) to generalize it for supporting other devices along sensors constituting the IoT, then also people involved and their activities related to those devices. Therefore, few new concepts are added or other ones reused from existing ontologies, described in following. *Instant* is the main class describing the time occurrence, which is reused from the Time[14] ontology. It has a specific data property, *hasTimestamp*, describing the exact time and date of the occurrence. Another concept, *Place*, is as well reused from DBpedia[15] and which through object property *closeTo* describes the vicinity of one location to another. Besides that, data properties *hasLatitude* and *hasLongitude*, describe the coordinates of the specific location.

[14] http://www.w3.org/2006/time#
[15] http://dbpedia.org/ontology/#

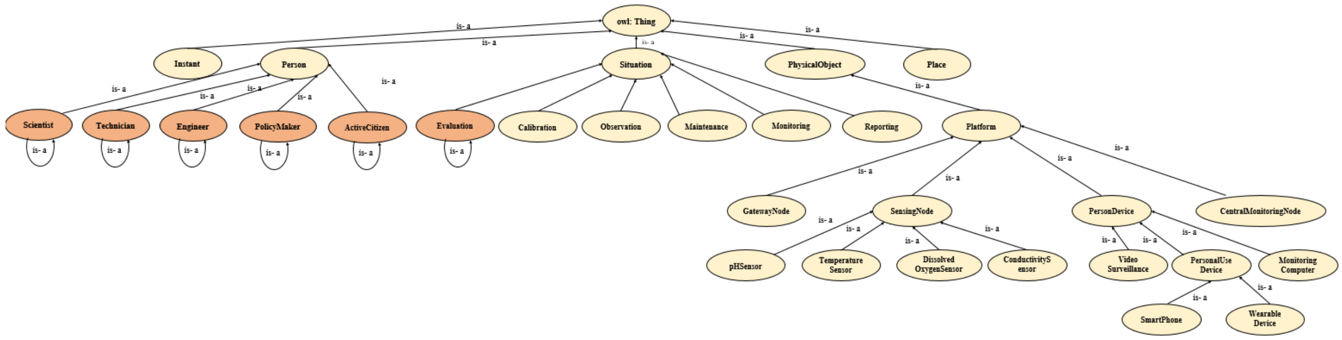

Figure 2. The DPAOntology – its subclass hierarchy

Modeling Devices in and around WSN

In the DPA ontology, one of the main parts of it deals with the description of devices of the WSN. As seen in Figure 3, the main class *PhysicalObject* represents the root. *PhysicalObject* is reused from SSN[16] ontology. Its sole subclass is:

- *Platform* – which is as well reused from the SSN ontology, also reused in INWS[17] ontology, represents the main entity, in which other components can be attached. Its subclasses are:
  - SensingNode – reused from the INWS ontology, represents all sensors attached to the system.
  - *CentralMonitoringNode* and *GatewayNode* – also reused from INWS. Sensors measure the specified parameters and send back data through the gateway node to the central monitoring node as the end point (a remote server) for further research.
  - *PersonDevice* - which is a new construct in our ontology, describes devices used by persons involved in the WSN. It has three direct subclasses: *MonitoringComputer* represents the device of the person that evaluates the data, while *PersonalUserDevice* represent the personal device of the person such as smartphone (SmartPhone class) or any portable wearable device (WearableDevice class). *VideoSurveillance* on in other hand represents the cameras attached to the WSN or any video report by the public, willing to contribute to the monitoring of water quality in general.

Properties of the *PhysicalObject* class are: *isCalibrated* which describes if the sensors are calibrated, *hasCamera* describing if the smartphone has a camera, as well as *hasDeviceID* and *hasDeviceStatus* identifying and describing the status of device respectively. The device ID represents in fact an RFID of the device in best case scenario under the assumption there are RFID tags available and manageable to attach to every singular device belonging to the input system.

[16] http://purl.oclc.org/NET/ssnx/ssn
[17] http://inwatersense.uni-pr.edu/ontologies/inws-core.owl

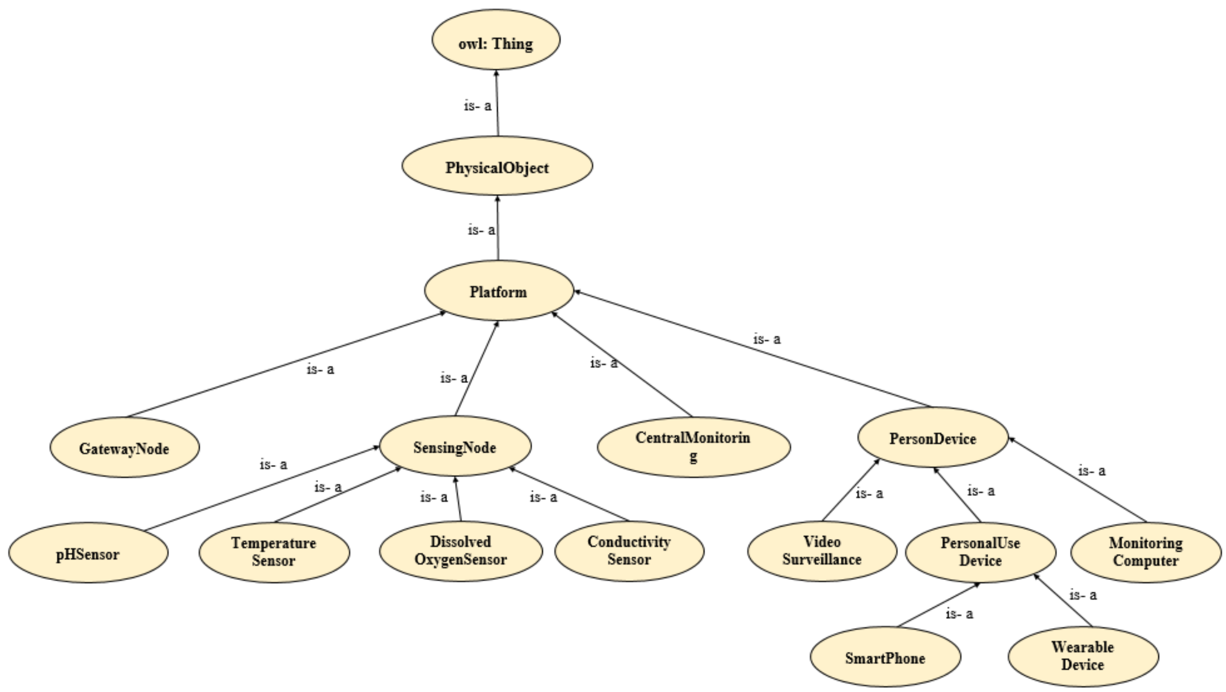

Figure 3. Modeling of devices – the subclass hierarchy

Modeling Person in and around WSN

Another important part of the ontology is person modeling, shown in Figure 4. The WSN captured involves a number of parties into its functionality. First of all, there is *Technician* directly involved with measuring required parameters. *Technician* is part of the internal structure of person involved and therefore described by the *internalPerson* object property. Another person part of the *internalPerson* property is *Engineer*, dealing with all the malfunctions of the system and more importantly the calibration of the sensors.

Other people involved, *ActiveCitizen*, *PolicyMaker* and *Scientist*, are all part of the external people involved with activities around the system and therefore described by the *externalPerson* property. *ActiveCizitizen* represents all people willing to contribute voluntarily to the water quality monitoring by sending video or picture of the possible pollution around the river. *PolicyMaker* represents other institutional people dealing with the water quality monitoring, which validates the data measured by sensors, whereas *Scientists* involve researchers that are willing to use data for their research.

All people represented in the ontology have a unique ID, described as data property *hasPersonID*, which distinguishes them from each other. That property is used as well as part of context inference, which let us understand if data can be released for public or not. If the data are validated, then the person ID would change to the ID of the *PolicyMaker*, therefore understanding that data are valid and can be used by other parties.

Besides that, there is another property involving *Person*, *hasActivity* property, which describes the situation that involves the person in.

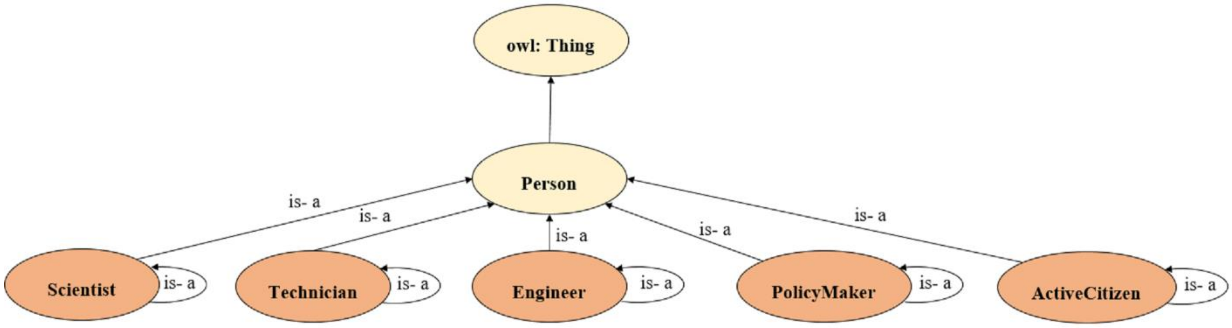

Figure 4. Modeling of person – the subclass hierarchy

Modeling Activities in and around WSN

Another part of the ontology describes the concepts related to the activities of the actors involved in the system. As seen in Figure 5, the top class in the activities part is *Situation*, which is reused from DUL[18] SSN ontology. Its subclasses are:

- *Calibration* – performed for the system and which models the calibration process.
- *Evaluation* – modeling the validation of the data.
- *Maintenance* – modeling the maintenance process of the system.
- *Monitoring* – modeling the monitoring of the system performance.
- *Observation* – reused from SSN ontology, with its properties *isMeasured* and *meassurementID.*
- *Reporting* – modeling a report from active citizens, with property *isReported.*

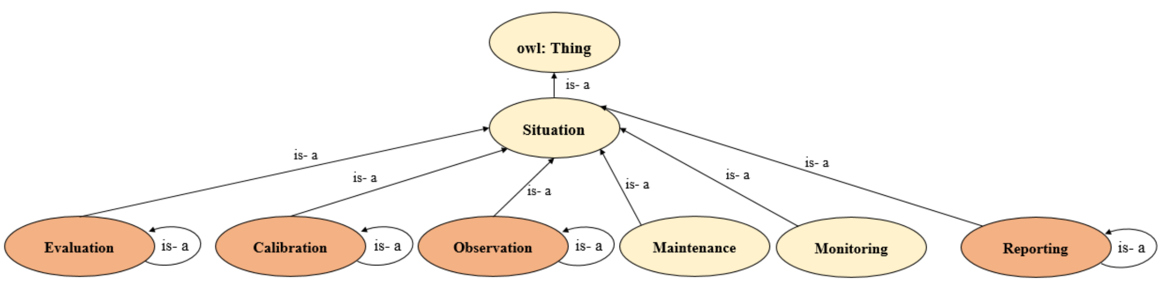

Figure 5. Modeling of activities – the subclass hierarchy

SEMANTIC MULTI-SERVICE WSN PLATFORM FOR THE IOT

The WSN platform as any other device constituting the IoT modeled through the DPA ontology intertwines several services, making thus the platform of the most IoT essence. First of all, by using the ontology it enables inferring contextual data, which are mined and provide more valuable insight then by using only raw data. Then, from another perspective, it allows for semantic WSN web enablement and further more linking WSN data through the URI. In the rest of this section, each service will be discussed separately in details accompanied with a possible scenario for each of them.

[18] http://www.loa-cnr.it/ontologies/DUL.owl#

Service 1: Contextual Data Mining over WSN Supported by Ontology

Data produced by sensors are accounted as raw data. Those direct sensor measurements are saved in a database for further analysis. Before the analysis, data has to be prepared and preprocessed for the process. This involves usage of filters, removal or substitution of missing data and discretizing or division of data into several bins. After that, by usage of state of the art algorithms, one will be able to extract valuable association rules.

Association rules describe the relation between several components. Those components can be raw data or even metadata – semantically enriched data. Even though association rules are used a lot in order to get those relations between raw data, only in the last few years the relations between metadata has been started to be drawn. That rose an interesting research area of semantically enriched metadata evaluation through association rules. Therefore, in our case, the process of the association rules over raw data is compared to the process of the association rules on metadata.

The same process applied on raw data will be identically conducted on the data extracted from ontologies. According to several authors (Abedjan, 2013, Bytyçi, 2016), semantically enriched data will provide further comprehension of the raw data. That enrichment is done through usage of ontologies used to describe concepts involving device, person, and activity, and their relations in the process. All of them are modelled into the ontology. By populating the ontology with data and then using a reasoner, further number of relations would be obtained. In Figure 6, a part of the process that results in rules obtained over semantic data is presented. As per case of raw data, data are preprocessed and prepared for the data mining process and in the end the resulted rules are obtained.

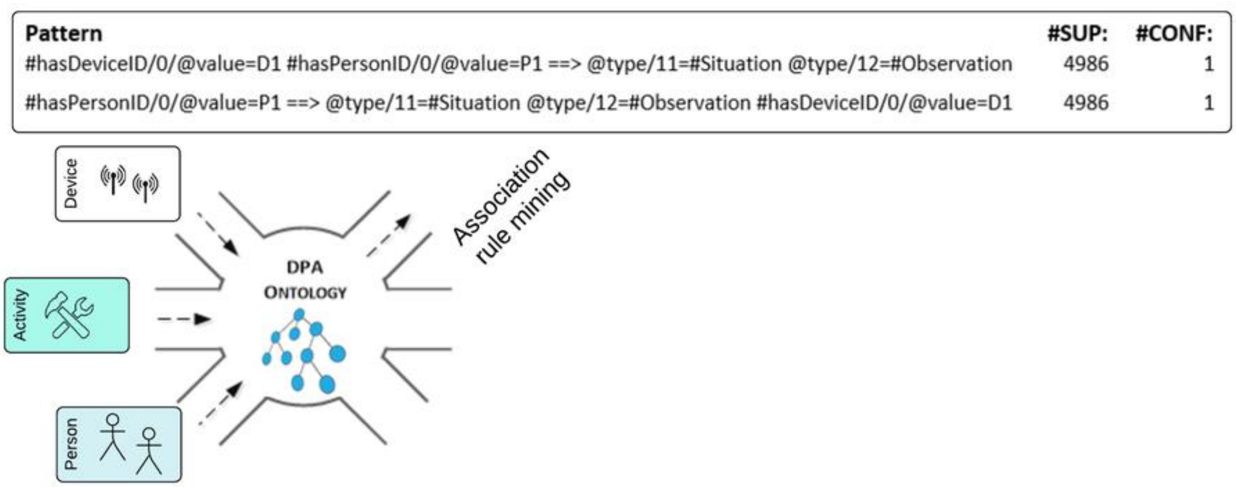

Figure 6. Association rules on semantic data

A possible case scenario from this approach would be knowing which information should be released for general public to perceive. Let's suppose that a technician, who does only the measurements, does the measurement with a mobile device equipped with sensors in a part of the river. Measurements will be sent to the database, together with technician ID, timestamp of the measurement and coordinates of the place. Now, according to the activity occurred, it can be concluded that a measurement is performed, which is obvious. But, in the other hand, when ID of the person changes to another value, for example from technician to engineer or someone else, depending on that, one can conclude if a validation of data has occurred or calibration has occurred or even if a reporting from somebody outside (an active citizen) has been performed.

In order to achieve that, the ontology is saved into appropriate format for further mining with association rules. Beforehand, as in the case of raw data, metadata will be preprocessed, where several supervised methods will be used such as normalization or discretization. Then, the Apriori algorithm (Agrawal, 1996) will be used over enriched sensor data. Results obtained are then compared to the results obtained by same Apriori algorithm in raw data.

Some of the raw data association rules, with the highest confidence, are presented in Table 1. There are only few strong relations derived. One of them, describes a strong relation between a person with a specific ID and a device with a specific ID. But, there is no other rules, not amongst the parameters itself or parameters and devices and people, indicating a relation. Of course, on the raw part of the data, there is no description of the activity that have occurred. The activity is derived as the context in the ontology, as a relation between person and device. For example, the measurement activity is reasoned from the person ID and device ID. So, if the person ID belongs to a technician and device ID to the equipment used for measurement, then it would be possible to infer which activity is performed. That would be of great benefit, since in the case of validated data, it would enable withholding the not validated data, and not let them release to the public.

Pattern	#SUP:	#CONF:
DeviceID=D1 ==> PersonID=P1	4991	1
PersonID=P1 ==> DeviceID=D1	4991	1
Temperature='\'(5.666667-10.333333]\'' DeviceID=D1 ==> PersonID=P1	1798	1
Temperature='\'(5.666667-10.333333]\'' PersonID=P1 ==> DeviceID=D1	1798	1

Table 1 Association rules on raw data

On the other hand, when association rules are applied on context ontology, an overwhelming number of rules are derived. For argument, we have presented in Table 2 only a couple of them, which in fact back up our initial claim – a strong relation between device-person-activity triple of concepts. It should be emphasized that the number of rules derived from the DPA context ontology data is significantly higher, i.e., more than 10 times the number of rules derived from applying the same algorithm but on raw data. The most interesting among derived rules is the one that creates a connection not only between person and device, but also the activity. Therefore, one can relate the person with the specific ID to the situation or activity conducted by him, through a specific device. This specific rule, acclaims the aid of the semantics by the usage of context in deriving new rules, strengthening even more the claim in a previous work (Bytyçi, 2016).

Pattern	#SUP:	#CONF:
#hasDeviceID/0/@value=D1 #hasPersonID/0/@value=P1 ==> @type/11=#Situation @type/12=#Observation	4986	1
#hasPersonID/0/@value=P1 ==> @type/11=#Situation @type/12=#Observation #hasDeviceID/0/@value=D1	4986	1

Table 2. Association rules on context ontology data

Service 2: Semantic WSN Web Enablement

Wireless Sensor Networks (WSNs) are part of many research areas lately, which can be attributed to the development of sensors in particular or even to the paradigm of the IoT in general. Furthermore, sensor networks are being enabled through the Sensor Web providing thereby solutions to the web enabled WSN (Rouached, 2012) (Udayakumar, 2012). The Open Geospatial Consortium (OGC) defines standardization for the Sensor Web named Sensor Web Enablement (SWE), which is divided into two parts (Echterhoff, 2011):

a. *SWE Information model* – is comprised of conceptual language encodings that permits sensor observations visibility on the Internet. The SWE information model includes the following specifications: *Observations & Measurements (O&M)* - Standard models and XML Schema for encoding observations and measurements from a sensor, both archived and real-time. *Sensor Model Language (SensorML)* - describes sensors systems and processes associated with sensor observations. *Transducer Model Language (TransducerML or TML)* - describes transducers and supporting real-time streaming of data to and from sensor systems.

b. *SWE Service model* - is a set of Web Service specifications that allow a client to search and find the required information. The SWE Service model includes the following specifications: *Sensor Observations Service (SOS)* - Standard web service interface for requesting, filtering, and retrieving observations and sensor system information. This is the intermediary between a client and an observation repository or near real-time sensor channel. *Sensor Planning Service (SPS)* - a web service interface for requesting user-driven acquisitions and observations. *Sensor Alert Service (SAS)* - a web service interface for publishing and subscribing to alerts from sensors. *Web Notification Services (WNS)* - a web service interface for asynchronous delivery of messages.

The goal of SWE is to enable all types of Web and/or Internet-accessible sensors, instruments, also imaging devices to be accessible and controllable via the Web where applicable. The vision is to define and approve the standards foundation for "plug-and-play" Web-based sensor networks. Usually a sensor location is a critical parameter for sensors on the Web, and OGC is the world's leading geospatial industry standards organization (Botts, 2011).

In this paper, the focus is on *Sensor Observations Service (SOS)*. The SOS standard is applicable to use cases in which sensor data need to be managed in an interoperable way. This standard defines a Web service interface which allows querying observations, sensor metadata, as well as representations of observed features (Bröring, 2012). SOS has three mandatory "core" operations (Na, 2007):

- *GetObservation* - provides access to sensor observations and measurement data via a spatio-temporal query that can be filtered by phenomena
- *DescribeSensor* - enables querying of metadata about the sensors and sensor systems available by an SOS server.
- *GetCapabilities* - provides access to metadata and detailed information about the operations available by an SOS server.

Following the SEMDPA architecture suggested, a prototype system in Java is developed, which implements standards like SWE, respectively version 2.0 of the SOS standard (SOS 2.0 relies on the OGC O&M) to encode data gathered by sensors (Bröring, 2012)). Figure 7 shows the system architecture covering this service as a valuable output of the SEMDPA architecture.

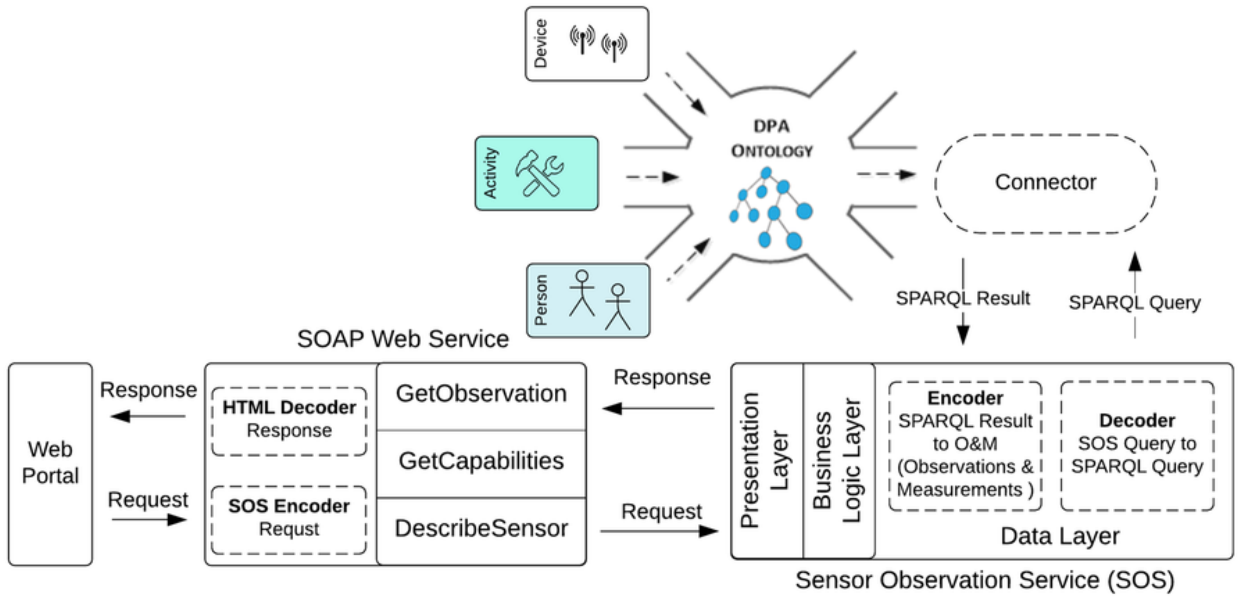

Figure 7. System Architecture of SEMDPA Web Services

In the following, the *GetObservation* operation is used to explain the functioning of the system stepwise. Given that the *GetObservation* function is the heart of the SOS (Henson, 2009), it is chosen among the three mandatory SOS operations to next describe the process of obtaining information about the sensor observations: Client posts a request to a web portal to read the measurements that have been conducted, specifying thereby optionally different filters such as: the time period within which measurements are made; phenomens like: temperature, electrical conductivity, pH, dissolved oxygen (DO), turbidity, biochemical oxygen demand (BOD), etc, measurement locations, sensors, etc. Such a request must be translated (encoded) into an SOS query of the *GetObservation* operation which is enabled through the Simple Object Access Protocol (SOAP) web service. The encoded client's request into SOS query is then transmitted to the SOS server. The SOS on Business Logic Layer makes the validation of the request. If the request is not valid then returns an exception report, otherwise forwards it to the Data Layer (the Decoder) which further decodes it, namely the SOS query into a SPARQL query. Depending on the filters that SOS query contains, the SPARQL query is generated dynamically applying these filters. The SPARQL query is executed over the ontology, in this case the DPA ontology, to extract required information (the measurements). It is worth mentioning that for the execution of the SPARQL query over the ontology, the Java library called Jena[19] Ontology API has been used. The result of the SPARQL query which is in format XML is converted (encoded) through the Encoder into *Observation & Measurement* (O&M) format of the OGC, because the response form *GetObservation* must be encoded in O&M. The response through SOS server conveys to the SOAP web services, in which through the HTML Decoder is done the decoding of the

[19] jena.apache.org/documentation/ontology/index.html

response from O&M format to HTML table format and is displayed to the client on the web portal.

An Example GetObservation Operation in SEMDAP

Next, an example demonstrating a *GetObservation* request and its SOS query encoded, as well as its corresponding SPARQL query and the *GetObservation* response is provided.

A. GetObservation request

An example request includes finding all the measurements made on locations Plemetin or Mitrovica, for phenomena such as temperature and electrical conductivity which have been measured from sensors Sensor1_Temp, Sensor2_Cond, Sensor3_Temp or Sensor4_Cond within the time period from 2016-01-19 14:00:00 and 2016-01-19 14:05:00. An abstract of the *GetObservation* request is presented below:

GetObservation(
(featureOfInterest := Plemetin OR Mitrovica) AND
(observedProperty := Temperature OR Conductivity) AND
(procedure := Sensor1_Temp OR Sensor2_Cond OR Sensor3_Temp OR Sensor4_Cond) AND
(temporalFilter := BETWEEN 2016-01-19T14:00:00 AND 2016-01-19T14:05:00.000))

B. SOS query

The request on *GetObservation* operation encoded into an SOS query is shown in Figure 8.

```
<sos:GetObservation xmlns:sos=... service="SOS" version="2.0.0"...>
  <sos:temporalFilter>
    <fes:During>
      <fes:ValueReference>phenomenonTime</fes:ValueReference>
      <gml:TimePeriod gml:id="tp_1">
        <gml:beginPosition>2016-01-19T14:00:00.000+01:00</gml:beginPosition>
        <gml:endPosition>2016-01-19T14:05:00.000+01:00</gml:endPosition>
      </gml:TimePeriod>
    </fes:During>
  </sos:temporalFilter>
<sos:featureOfInterest>http://inwatersense.uni-pr.edu/ontologies/inws-
core.owl#Plemetin</sos:featureOfInterest>
<sos:featureOfInterest>http://inwatersense.uni-pr.edu/ontologies/inws-
core.owl#Mitrovica</sos:featureOfInterest>
<sos:observedProperty>http://inwatersense.uni-pr.edu/ontologies/inws-
core.owl#Temperature</sos:observedProperty>
<sos:observedProperty>http://inwatersense.uni-pr.edu/ontologies/inws-
core.owl#Conductivity</sos:observedProperty>
<sos:procedure>http://inwatersense.uni-pr.edu/ontologies/inws-
core.owl#Sensor1_Temp</sos:procedure>
<sos:procedure>http://inwatersense.uni-pr.edu/ontologies/inws-
core.owl#Sensor2_Cond</sos:procedure>
<sos:procedure>http://inwatersense.uni-pr.edu/ontologies/inws-
core.owl#Sensor3_Temp</sos:procedure>
<sos:procedure>http://inwatersense.uni-pr.edu/ontologies/inws-
core.owl#Sensor4_Cond</sos:procedure>
  <sos:responseFormat>http://www.opengis.net/om/2.0</sos:responseFormat>
</sos:GetObservation>
```

Figure 8. Example GetObservation Request as SOS query

The given *GetObservation* request, its SOS query, contains following properties as defined in (Bröring, 2012) (Cox, 2011):

- *temporalFilter* - specifies a filter for a time property of requested observations.
- *featureOfInterest* - pointer to a feature of interest for which observations are requested.
- *observedProperty* - pointer to an observedProperty for which observations are requested.
- *procedure* - pointer to a procedure for which observations are requested. It defines a filter for the procedure property of the observations.
- *responseFormat* - identifier of desired responseFormat for the requested observations (Default is O&M 2.0 [OGC 10-004r3/ISO 19156] identified by the value http://www.opengis.net/om/2.0).

C. SPARQL query

The *Decoder SOS Query to SPARQL Query* component of the proposed system will decode the request from *GetObservation* SOS queryinto SPARQL query, depicted in Figure 9, which will then be executed over the DPA ontology to generate the required result.

```
PREFIX ...
SELECT ?sos ?hasTimestamp ?hasConductivity ?hasTemperature ?hasLocation
WHERE {
?sos inws:hasTimestamp ?hasTimestamp.
?sos inws:hasTemperature ?hasTemperature.
?sos inws:hasConductivity ?hasConductivity.
?sos inws:hasLocation ?hasLocation.
?sos inws:hasDeviceID ?hasDeviceID.
FILTER ( ?hasTimestamp >= "2016-01-19T19:14:00"^^xsd:dateTime &&
         ?hasTimestamp <= "2016-01-19T19:14:05"^^xsd:dateTime ).
FILTER (?hasLocation = "Mitrovica" || ?hasLocation = "Plemetin").
FILTER (?hasDeviceID = "Sensor1_Temp" || ?hasDeviceID = "Sensor2_Cond" ||
        ?hasDeviceID = "Sensor3_Temp" || ?hasDeviceID = "Sensor4_Cond" ).
}
```

Figure 9. Example SPARQL Query

D. GetObservation response

Encoder SPARQL Result to O&M (Observation & Measurements) will encode the result of SPARQL in the O&M standard because the response format of *GetObservation* request (SOS 2.0) must be encoded according (Cox, 2011) to O&M. Figure 10 shows an excerpt of the *GetObservation* response.

```
<sos:GetObservationResponse ...>
  <observationData>
    <om:OM_Observation gml:id="o1">
      <om:type xlink:href="http://www.opengis.net/def/observationType/OGC-OM/2.0/OM_Measurement"/>
      <om:phenomenonTime>
        <gml:TimeInstant gml:id="phenomenonTime_1">
          <gml:timePosition>2016-01-19T14:00:00.000+01:00</gml:timePosition>
        </gml:TimeInstant>
      </om:phenomenonTime>
      <om:resultTime xlink:href="#phenomenonTime_1"/>
      <om:procedure xlink:href="http://inwatersense.uni-pr.edu/ontologies/inws-
core.owl#Sensor1_Temp"/>
      <om:observedProperty xlink:href="http://inwatersense.uni-pr.edu/ontologies/inws-
core.owl#Temperature"/>
      <om:featureOfInterest xlink:href="http://inwatersense.uni-pr.edu/ontologies/inws-
core.owl#Mitrovica"/>
      <om:result xsi:type="gml:MeasureType" uom="C">15.3</om:result>
    </om:OM_Observation>
  </observationData>
      ...
</sos:GetObservationResponse>
```

Figure 10. Example GetObservation response

Service 3: Linked Open WSN Data

As part of the lifecycle of Linked Open Data (LOD), there are three main phases required for creating and publishing data following this paradigm, which will in this section be discussed in the context of our SEMDPA architecture. Initially, when talking about generating the LOD, it should be noted that in our case those data are generated by sensor measurements, as presented in Figure 11.

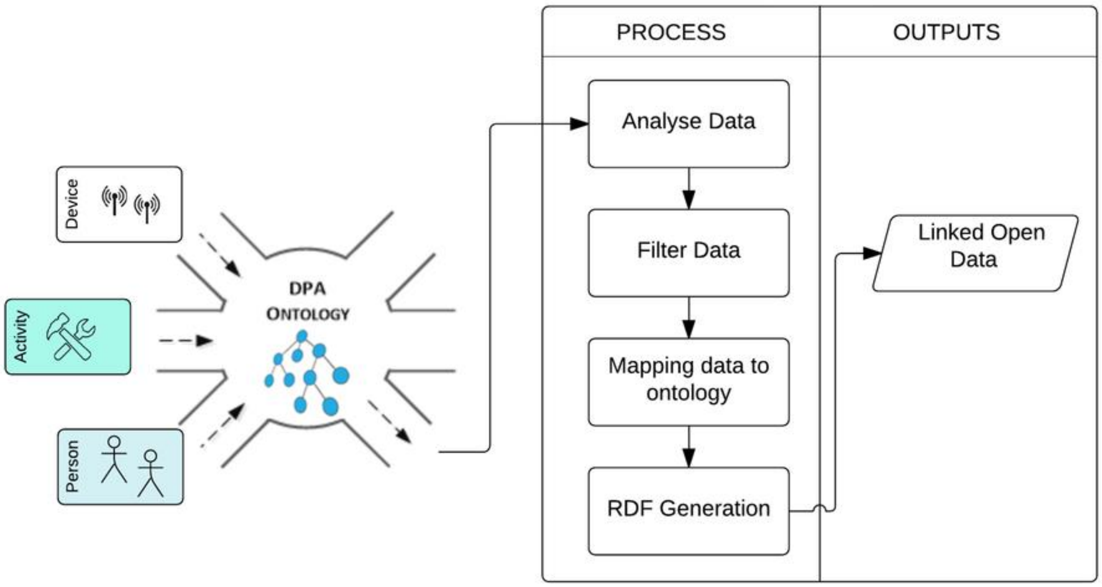

Figure 7. Publishing process of sensor data as Linked Open Data

The initial step of creating the LOD includes several methods. One of the initial methods is analysis of data. Since sensors generate a lot of data with some being also not relevant for publishing as Linked Open Data, one needs to analyze the data and filter the ones suitable for the process of creation of LOD. Thus it is crucial to know the content of data generated from sensors. After that, as part of filtering, parts of dataset to be published as LOD are extracted.

The next phase after data analysis and filtering is the mapping of data to an ontology. It incorporates our sensor architecture, which in addition related to the DPA ontology is presented in Figure 12. In this phase, it is of imperative importance that a proper ontology is being used, involving concepts and relations between data.

```
map:sensorvalues_Temperature a d2rq:PropertyBridge;
d2rq:belongsToClassMap map:sensorvalues;
d2rq:property vocab:sensorvalues_Temperature;
d2rq:propertyDefinitionLabel "sensorvalues Temperature";
d2rq:column "sensorvalues.Temperature";
d2rq:datatype xsd:integer;
```

Figure 812. Mapping data example

As it is requirement of the Semantic Web community, the ontology should either extend one or more existing ontologies, or be defined from the scratch in cases the structure of data is not fully supported by the existing ontologies. The mapping schema between data and ontology can be created by linking data to respective classes in ontology or can be generated automatically by usage of a specific tool. An example of such a mapping tool is D2RQ[20] generate-mapping tool, which creates a D2RQ mapping file by analyzing the schema of an existing database (some graphical details are shown in Figure 13). Created mapping file, maps each table of the database to a new RDFS class that is based on the table's name, and maps each column to a property based on the column's name. In our case mapping is created manually according to DPA ontology (example presented in Figure 12), even though this mapping file can be used as it is or can be customized, as needed (Bizer, 2004).

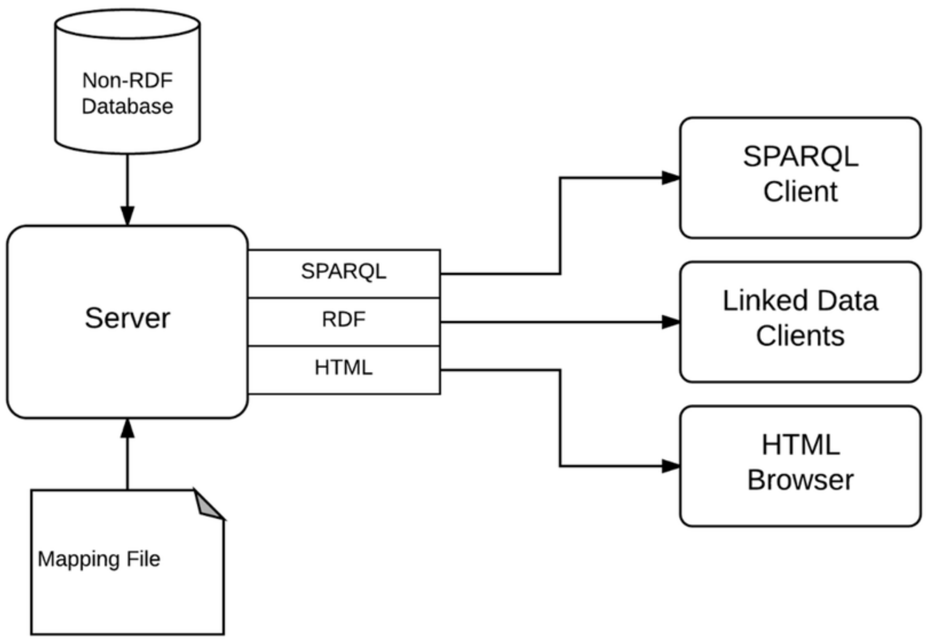

Figure 13. Server Architecture

The next phase in the process of creating LOD is conversion of data to LOD. Even though several approaches for converting data to LOD are mentioned, they might be grouped in two main approaches: generating RDF and using RDF Storage to access them, or on the fly conversation. In the first approach, LOD are converted from sensor data and stored in a triple store which can be queried via SPARQL queries as shown in Figure 9. This is the approach that we have used. In Figure 9, the SPARQL example is shown, used for retrieving all sensor measurements for specified location, time and device. In the second approach, sensor data are stored in database and converted in RDF on the fly when they requested. An example of it would be the D2RQ-query tool that allows executing SPARQL queries against a D2RQ-mapped relational database from the command line. This can be done with or without a D2RQ mapping file. If a mapping file is specified, then the tool will query the virtual RDF graph defined by the mapping and if no mapping file is

[20] d2rq.org

specified, then the tool will use the default mapping in D2RQ to generate-mapping for the translation (Bizer, 2004).

For the different components of SEMDPA model, Water Sensor Observation Service requires URIs as link of observations to the LOD. Those URIs are assigned to the components such as sensorvalues[21] by appending the component type such us Conductivity[22]to the URI identifying the authority. The Component sensorvalues[22] refers links to all sensor descriptions. Consequently, Conductivity[23] refers the description of a conductivity sensor and links to the produced observations.

An example showing the relation of a person to LOD is when the involved person authenticates himself through his/her Facebook account which has a specific URI (for example arten.avdiu.3[23]), which may link to other relevant information about the person. Also, for each involved device, a URI containing RFID tag[24] is generated, which as well may link to other relevant information about the device.

For each activity of the actors involved in the system is similarly assigned an URI, such as Situation[25], or a specific feature may be accessed by appending identifiers of those resources to the base URI. For example, the reference Reporting[26] points to all reports made by the person[26].

DISCUSSION AND FUTURE WORK

Even though the IoT was introduced initially to connect RFID device information to the internet, it has evolved through the years. Now, the IoT is envisioned as an infrastructure that intelligently links living and non-living things to the Internet, creating enormous amount of useful data through their intercommunication to our everyday life. In line with its aim, our proposed SEMDPA architecture enables the triangular ontological relation between device, person, and activity conducted by both. With the help of web semantics as the center of the architecture, and by usage of new evolving technologies such as context-aware data mining techniques, semantic-enriched sensor web enablement standards and linked data methods, our approach achieves the interlinking of the previous under the umbrella of IoT. Although the architecture suggested is illustrated to work in the domain of WSN data at the input, it is generic as to capture the semantics of input sensor or even other devices' data from other domains, but still similarly gain due to services provided at the output. This denotes the most contribution of this paper, with a generic architecture and the triangle DPA ontology in line with the IoT vision in the middle as a crossroad towards multitude of useful services at the output. The SEMDPA architecture is moreover distinguished for its ease of service provision via a single Web, and for the richness of service provision due to semantics encapsulated by mean of ontology in the middle of the architecture.

21 http://inwatersense.uni-pr.edu/data/sensorvalues
22 http://inwatersense.uni-pr.edu/data/sensorvalues#Conductivity
23 https://web.facebook.com/arten.avdiu.3
24 http://inwatersense.uni-pr.edu/vocab/sensorvalues_DeviceID/334bte3g5
25 http://inwatersense.uni-pr.edu/data/Situation
26 http://inwatersense.uni-pr.edu/data/Situation/Reporting

It is achieved, by using data mining techniques, to find hidden relation between data, previously coated by the semantics of the given context. Furthermore, by using SWE, or in case of devices other than sensors at the input, their corresponding Web-enabled standards, it creates a peculiar prototype to connect the sensors, or certain other devices, via a semantic web. And in the end, by integrating them to linked data, it facilitates for further global querying and analyzing.

Despite previous work existing in provision of either of these services alone or in certain twin combinations as discussed in the related work, none has achieved to present a generic lightweight architecture modeling the triple device-person-activity of actors at the input, and a *triple of services or even more* at the output benefiting from such an architecture.

In the future, with aim on evolving the architecture, each component will be further supplemented to reflect the peculiarities of the certain domain and devices at the input, as well as services acquired at the output in the context of IoT

REFERENCES

52North Sensor Web Community (2004). Retrieved 12.11.2016 from http://52north.org

Abedjan, Z., & Naumann, F. (2013). Improving rdf data through association rule mining. *Datenbank-Spektrum*, *13*(2), 111-120.

Aggarwal, C. C., Ashish, N., & Sheth, A. (2013). The internet of things: A survey from the data-centric perspective. In *Managing and mining sensor data* (pp. 383-428). Springer US.

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast discovery of association rules. *Advances in knowledge discovery and data mining*, *12*(1), 307-328.

Ahmedi, L., Jajaga, E., & Ahmedi, F. (2013, October). An ontology framework for water quality management. In *Proceedings of the 6th International Conference on Semantic Sensor Networks-Volume 1063* (pp. 35-50). CEUR-WS. org.

Ahmedi, L., Sejdiu, B., Bytyçi, E., & Ahmedi, F. (2015). An Integrated Web Portal for Water Quality Monitoring through Wireless Sensor Networks. *International Journal of Web Portals (IJWP)*, *7*(1), 28-46.

Ashton, K. (2009). That ‘internet of things’ thing. *RFiD Journal*, *22*(7), 97-114.

Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks*, *54*(15), 2787-2805.

Berners-Lee, T. (2010). Linked Data Design Issues (July 2006). *URL= http://www. w3. org/DesignIssues/LinkedData. html, última visita junio de*.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, *284*(5), 28-37.

Bizer, C., & Seaborne, A. (2004, November). D2RQ-treating non-RDF databases as virtual RDF graphs. In *Proceedings of the 3rd international semantic web conference* (ISWC2004) (Vol. 2004). Springer.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, 205-227.

Botts, M., Percivall, G., Reed, C., & Davidson, J. (2006, October). OGC® sensor web enablement: Overview and high level architecture. In *International conference on GeoSensor Networks* (pp. 175-190). Springer Berlin Heidelberg.

Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., Liang, S., Lemmens, R. (2011). New generation sensor web enablement. *Sensors*, *11*(3), 2652-2699.

Bröring, A., Stasch, C., & Echterhoff, J. (2012). OGC sensor observation service interface standard. *Open Geospatial Consortium Interface Standard*, 12-006.

Bytyçi, E., Ahmedi, L., & Kurti, A. (2016). Association Rule Mining with Context Ontologies: An Application to Mobile Sensing of Water Quality. In *Metadata and Semantics Research: 10th International Conference*, MTSR 2016, Göttingen, Germany, November 22-25, 2016, Proceedings (pp. 67-78). Springer International Publishing.

Calbimonte, J. P., Jeung, H., Corcho, O., & Aberer, K. (2011, October). Semantic sensor data search in a large-scale federated sensor network. In *Proceedings of the 4th International Conference on Semantic Sensor Networks*-Volume 839 (pp. 23-38). CEUR-WS. org.

Compton, M., Barnaghi, P., Bermudez, L., GarcíA-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., & Huang, V. (2012). The SSN ontology of the W3C semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web*, *17*, 25-32.

Conover, H., Goodman, H. M., Zavodsky, B., Regner, K., Maskey, M., Lu, J., Li, X., Botts, M. & Berthiau, G. (2008). Intelligent Assimilation of Satellite Data into a Forecast Model Using Sensor Web Processes and Protocols. In *Earth Science Technology Conference*.

Cox, S. (2011). Observations and measurements-xml implementation. OGC document.

Cygankiak, R. World Wide Web Consortium. (2014). Best practices for publishing linked data.

Devaraju, A., & Kauppinen, T. (2012). Sensors tell more than they sense: Modeling and reasoning about sensor observations for understanding weather events. *International Journal of Sensors Wireless Communications and Control*, *2*(1), 14-26.

Echterhoff J. (2011) Open Geospatial Consortium. OpenGIS SWE Service Model Implementation Standard. *Open Geospatial Consoritum*, 155.

Giusto, D., Iera, A., Morabito, G., & Atzori, L. (Eds.). (2010). The internet of things: 20th Tyrrhenian workshop on digital communications. Springer Science & Business Media.

Henson, C. A., Pschorr, J. K., Sheth, A. P., & Thirunarayan, K. (2009, May). SemSOS: Semantic sensor observation service. In *Collaborative Technologies and Systems*, 2009. CTS'09. International Symposium on (pp. 44-53). IEEE.

Hyland, B., & Wood, D. (2011). The joy of data-a cookbook for publishing linked government data on the web. In *Linking government data* (pp. 3-26). Springer New York.

Jajaga, E., Ahmedi, L. & Ahmedi, F. (2017a). StreamJess: Stream Data Reasoning System for Water Quality Monitoring. International Journal of Metadata, Semantics and Ontologies – In press.

Jajaga, E. & Ahmedi, L. (2017b). C-SWRL: SWRL for Reasoning over Stream Data information. SDI Workshop, IEEE ICSC 2017.

Jajaga, E., Ahmedi, L. and Ahmedi, F. (2016). StreamJess: Enabling Jess for Stream Data Reasoning and the Water Domain Case (Demo paper) 20th International Conference on Knowledge Engineering and Knowledge Management (EKAW2016), Bologna.

Keßler, C., & Janowicz, K. (2010, November). Linking sensor data-why, to what, and how? In *Proceedings of the 3rd International Conference on Semantic Sensor Networks*-Volume 668 (pp. 48-63). CEUR-WS. org.

Lee, Y. J., Trevathan, J., Atkinson, I., & Read, W. (2015). The Integration, Analysis and Visualization of Sensor Data from Dispersed Wireless Sensor Network Systems Using the SWE Framework. *Journal of Telecommunications and Information Technology*, (4), 86.

Lefort, L., Henson, C., Taylor, K., Barnaghi, P., Compton, M., Corcho, O., O., Garcia-Castro, R., Graybeal, J., Herzog, A., Janowicz, K. & Neuhaus, H. (2011). Semantic sensor network xg final report. *W3C Incubator Group Report*, *28*.

Le-Phuoc, D., Quoc, H. N. M., Parreira, J. X., & Hauswirth, M. (2011). The linked sensor middleware–connecting the real world and the semantic web. *Proceedings of the Semantic Web Challenge*, *152*.

Na, A., & Priest, M. (2007). Sensor observation service. *Implementation Standard OGC*.

Perera, C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2014). Context aware computing for the internet of things: A survey. *IEEE Communications Surveys & Tutorials*, *16*(1), 414-454.

Rouached, M., Baccar, S., & Abid, M. (2012, June). RESTful sensor web enablement services for wireless sensor networks. In *2012 IEEE Eighth World Congress on Services* (pp. 65-72). IEEE.

Schmachtenberg, M., Bizer, C., Jentzsch, A., & Cyganiak, R. (2014). Linking open data cloud diagram 2014.

Sheth, A., Henson, C., & Sahoo, S. S. (2008). Semantic sensor web. IEEE Internet computing, 12(4), 78-83.

Spahiu N., Ahmedi L. & Bouju A. Implementation of sensor observation services to monitor moving objects. A *master thesis* under the joint supervision of Ahmedi L., Bouju A. University of Prishtina, Kosovo, and University of La Rochelle, France, 2010.

Udayakumar, P., & Indhumathi, M. (2012). Semantic web based Sensor Planning Services (SPS) for Sensor Web Enablement (SWE). arXiv preprint arXiv:1207.5310.