Seventh International Conference on Information Science and Technology
Da Nang, Vietnam; April 16-19, 2017

# A Semantic Data Model For the Interpretion of Environmental Streaming Data

Truong Khanh Duy[1,2], Gerald Quirchmayr[1]
Faculty of Computer Science, University of Vienna
*Vienna, Austria*

Amin Tjoa[3] , Hoang Huu Hanh[2] University of Technology *(Vienna, Austria)*; - Hue University *(Thua Thien Hue, Viet Nam)*

***Abstract*— IoT Data and smart data is the most important key in research on particulate domain. This paper we present a semantic data model for environment system by propose a light-weight semantic ontology approach for represent IoT data with Semantic Web. By using a cross-domain knowledge by extend SSN ontology combine with other domain specific to mapping sensor streaming data and annotation sensor.**

***Keywords—semantic web, stream data, RDF stream, SSN, light -weightoOntology, smart data.***

## I. INTRODUCTION

Nowadays, we are living in a connected-world, Internet, information technology and telecommunications are being developing and become an indispensable part of everyday life. Sensors and sensor network are being deployed in our environment system for multi purposes. Every second, massive of data are being produced by sensors. Mobile devices, personal computer are publishing data on the Internet. More sensors, RFID tags, software components provided information as data streams and it using for some fields of application: traffic management, health monitor, financial trading. The decision support system is processing and analyze stream data to diagnose the actual state of a system allowing adequate reactions on critical situations.

The heterogeneity of data as well as environments sensors is a key on sensors world. Many type of sensor network deployments for represent the information [1]. The data models and schema is different (wind speed, air pollution, weather) and complexity of devices make seamless integration of data and service. For the higher-level of applications and service, data must be archive and annotation, a solution [2] [3] to tackle this issue by using service-oriented to provide a common interface that represent IoT devices and data.

The Sensor Web Enablement group has created a XML standard model for sensor network called SensorML [4] to descripted sensor systems. The XML descript in SensorML provide means and link from other domain knowledge in limited. This model represented in RDF and enable linking and annotating data but this model is complex and not appropriate for large-scale data processing. SSN ontology [5] provide a basic description and observation for sensors semantic networks by the W3C Semantic Sensor Network Incubator group, however it does not provide detailed descriptions for measurement data and do not have any basic reasoning for development high level IoT applications. For understanding the business meaning of stream data, data must be enriched with semantically background knowledge. We propose a solution to deal with heterogeneity by apply ontology base approach in this article. We summarize our contributions as follows:

- *Using SSN ontology with cross-domain knowledge by design a light-weight ontology solution for annotation and present sensors and sensor data.*
- *Proposal RDF stream modal for store sensors stream data.*

The second chapter is expanding the requirements of environment sensor data and the keys of streaming data for query and reasoning. Afterwards, we discuss the possible data model and present conceptual modal. Final we'll show the framework and elementary results.

## II. REQUIREMENTS OF THE DATA MODEL

Sensors data are heterogeneous and variate dimensions like date type, stream, covered domain. To define a unifying data model, we must describe data properties and design an ontology base approach to convert incoming data. To reduce the amount of information transmitted over the network, data model must be lightweight and meet below requirements:

- *Data annotation:* the process incoming data can be seen as the data annotation processing. Raw data must be annotating by semantic description and defined by combined ontology.
- *Sensors, station information:* Along with sensor data, Sensor metadata must be store with stream data such as location, name, observation type etc.
- *Lightweight data model:* For handle big amount sensor data, lightweight model is very important. It helps easy to use, processing, store and query.

- *Reuse domain knowledge:* The vocabulary should be reusable and designed to be generic for easily pickup and combine with other data resource.

- *Semantic reasoning:* From the data with rich semantic description, it can be queried by other semantic web technology for semantic applications and services.

And besides that, environment data stream has other attributes such as size, location de-pendency, time dependency, life span etc. Using semantic description can provide machine-interpretable for environment data stream.

## III. Possible Technological

As shown above, a possible technology must be meet the requirements. The semantic descriptions of data stream follow a linked-data approach. Data must be defined by ontologies and link to exiting concept and store sensor metadata. Other hand, data model must be lightweight and easy to query and processing over semantic queries technology.

### A. JSON-LD

Current web technology has defined widely-used mechanisms such as REST APIs, HTML5, JSON and more backend technology to provide a web base sharing information. For share the meaning of data, the semantic web community has investigated for years to create a flexible well-defined model including many need aspects such as a powerful triple model and use URLs for type definitions. But these effort has occurred because non-wide user well known since some technologies are verbose and complicated with some discussions tend to be obscure.

A simple approach to JSON named values with triples in semantic web modal can be considered a breakthrough. JSON-LD *(JavaScript Object Notation for Linking Data)* [6] approach hides the entire semantic web mechanism from web developers, yet enables the full power of the semantic web to be used behind the scenes when in-formation is processed. The following show a richer model behind an JSON syntax [7]:

```
<script type="application/ld+json">
{
   "@context": {
   "foaf": "http://xmlns.com/foaf/0.1/",
   "title": "foaf:title",
   "name": "foaf:name",
                "homepage": {
   "@id":"foaf:workplaceHomepage",
   "@type": "@id"
                }
   },
   "@id": "http://me.markus-lanthaler.com",
   "@type": "foaf:Person",
                "title": [
                         {"@value":"Dipl.Ing.",
"@language":"de"},
                         {"@value":"MSc",
"@language": "en"}
              ],
       "name": "Markus Lanthaler",
       "homepage": "http://www.tugraz.at/" 19
   }
</script>
```

Fig. 1. Sample JSON FOAF Profile

This means that the semantic layer above need not mean hard to understand and it very friendly with developer. JSON-LD properties take the developers easy to building a Linked-data interchange application with JavaScript language. Additions, it familiar with current unstructured databases such as CouchDB or MongoDB and Web service.

### B. RDF-Stream

Semantic web technologies such as RDF and SPARQL have been applied for data stream over years, in what broadly call Linked Data Streams [8]. *"RDF is a standard model for data interchange on the web"* (http://www.w3.org/RDF/). RDF is a graph base data model which provides a grammar for its syntax. The newest RDF version (1.1), RDF syntax can be written in various concrete formats which called RDF serialization formats: Turtle, N-Triples, TriG, N-Quads and JSON-LD.

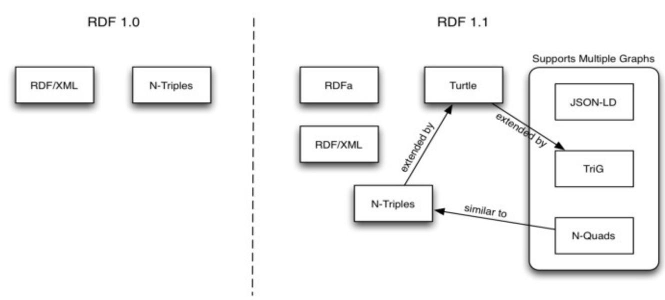


Fig. 2. RDF 1.1 serialization formats *(https://www.w3.org/TR/rdf11-new/)*

Above figure show JSON-LD is a concrete RDF syntax. Hence, a JSON-LD is an RDF document and a JSON and correspondingly represents an instance of an RDF model. Both RDF model and JSON-LD are Linked-data recommendation format for public semantic data. JSON-LD is appropriate for a better web API but it limit for semantic activity like reasoning, expressive query tools. Based on data model requirements, JSON-LD is suitable and already lightweight for linked-data model but for advance semantic query and stream processing, JSON-LD not mature yet. In our approach, we choose RDF data model for representing data and future semantic base activity.

## IV. Conceptual Data Modal

We propose a framework that combine SSN ontology with cross-domain knowledge to provide a common description for sensors data. Sensors often produce raw data and unstructured stream. On the other hand, they have different schema although they measure same phenomena values such as air pollution, wind speed. We used SSN ontology main concept and connected with other domain knowledge bellow described the multi domain ontologies and easily extend the data model.

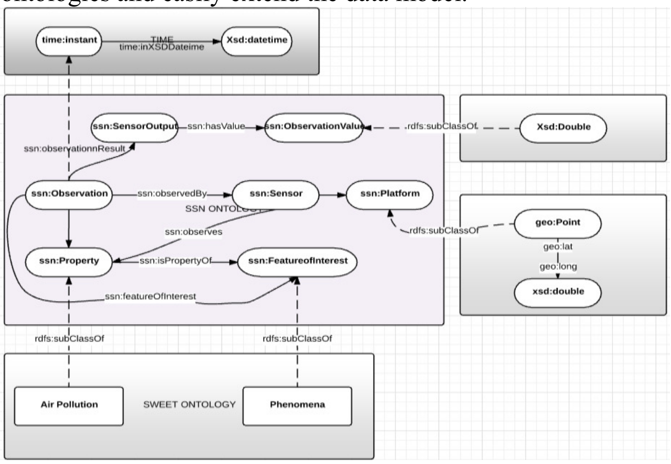


Fig. 3. Concept of the data model

Above figure show Data model concept that SSN ontology have combined with other domain knowledge to represent an Air quality: PM10 observation. It not only sensor observation value but also GEO location, date time information. In the next chapter we'll show the prototype of the concept that showed above. Mapping to RDF is an important phrase in our framework, to automatically mapping from raw data, the framework must be generating a mapping template according the ontologies. Using R2RML [9] Language, we indicate the Object, Subject and Predicate of RDF triple and mapping to RDF as follows:

- Every RDF stream will have subject map that identified by a URI.
- Subject is an instance of **SSN ontology observation**.
- Object is value of a SSN observation base on sensor model.
- Predicate will be an instance of ontology combined with SSN that descripted sensor data type.

We use PHP to provide frontend service and building framework. By provide a flexible RESTful web service, our framework provided a multi model output that all rich semantic description.

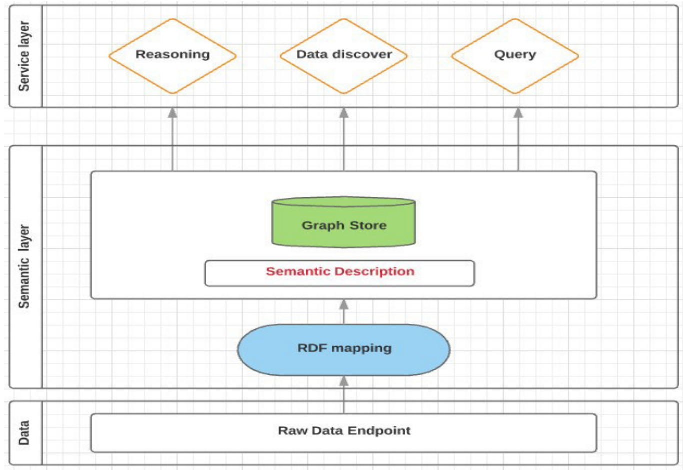


Fig. 4. Prototype of framework

Above prototype implemented by open source PHP and libraries. PHP backend process Data and store both triple store and RD for frontend service. We use Redstore – a lightweight RDF graph store using Redland library for store all RDF triple and support latest SPARQL 1.1. Graph store service is deployed on a dedicated server for stable and using SPARQL 1.1 HTTP [10] protocol for manage triple on our framework.

## V. IMPLEMENTATION STRATEGY AND PROTOTYPE

In this section we present a IoT web service URIs to identify sensor and stream data. We designed a human-friendly URI for represent data and easy for query. URI for station, sensor metadata. The URI will include sensor name and location for user friendly design.

http://itoserivce.hueuni.edu.vn/data/station/{idname}

For instance, below is RDF description Da Nang fixed station location on 41 Le Duan. Str, Da Nang, Viet Nam.

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-
schema#> .
@prefix ns0: <http://www.w3.org/ns/ssn#> .
@prefix geo:
<http://www.w3.org/2003/01/geo/wgs84_pos#> .

<http://iotservice.hueuni.edu.vn/data/station/DN
1>
  a <http://www.w3.org/ns/ssn#Observation> ;
  rdfs:Label "DN1 station" ;
  ns0:observedBy [
    a ns0:Sensor ;
    ns0:onPlatform [
      a ns0:Platform ;
      geo:Geometry [
        a geo:Point ;
        geo:lat 1.605440e+1 ;
```

```
        geo:long 1.082022e+2
      ]
    ] ;
    ns0:observedProperty _:genid2
  ] ;
  ns0:observes _:genid2 ;
  ns0:featureOfInterest [
    a ns0:FeatureOfInterest ;
    ns0:hasProperty _:genid2 ;
    rdfs:subClassOf
<http://sweet.jpl.nasa.gov/2.3/phenAtmo.owl#Atmo
sphericPhenomena>
  ] .

_:genid2
  a ns0:Property ;
  rdfs:subClassOf
<http://sweet.jpl.nasa.gov/2.3/phenEnvirImpact.o
wl#Pollution> .
```

Fig. 5. URI of Da Nang fixed atumatic station http://iotservice.hueuni.edu.vn/data/DN1

We can see the **ssn:observation** class that described a station, the **ssn:property** indicates the property of sensor and station (air pollution measurement) by SWEET ontology[11]. Furthermore, it shows the sensor's location, name, feature of interest that descripted by other cross-domain ontologies.
For Sensor's streaming data we using timestamp to identify data *http://iotservice.hueuni.edu.vn/data/sensor/{sensor_idname}/{timestamp}*

**Timestamp** is UNIX timestamp that descripted the update-time of sensor in a station. For an instance, we have air quality input of a fixed station in Da Nang, Viet Nam measures PM10, NOx, CO, O3 in semi real-time push:

TABLE I. SAMPLE AIR QUALITY MEASURE[a]

| Time | Observation | | | |
|---|---|---|---|---|
| | ***PM10*** | ***CO*** | ***NOx*** | ***O3*** |
| 17:50:00 | 1.52 | 2014.17 | 0.00 | 31.44 |
| 17:45:00 | 1.75 | 2239.17 | 0.00 | 32.75 |
| 17:40:00 | 1.59 | 1620.00 | 0.00 | 30.61 |
| 17:35:00 | 1.44 | 1783.79 | 0.00 | 28.52 |

a. Da Nang City in 28.03.2016 *(cem.gov.vn)*

We assume that the sensor produces raw and unstructured data streams, different schemas, although they both measure Air quality information: O3, NOx, CO, PM-10. Below is a sample RDF data that describe a sensor's data PM10 of fixed station. The **ssn:observedBy** indicates the property of the Station showed above figure 06. So forth for other measurement value O3, NO, NOx.

```
@prefix ssn: <http://www.w3.org/ns/ssn#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-
schema#> .
@prefix qudt: <http://qudt.org/schema/qudt#> .
@prefix time: <http://www.w3.org/2006/time#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>
.

<http://iotservice.hueuni.edu.vn/DN1/PM10_147696
6847>
  a ssn:Observation ;
  rdfs:Label "DN1 PM10" ;
  ssn:observationResult [
    a ssn:SensorOutput ;
    ssn:hasValue [
      a ssn:ObservationValue ;
      qudt:numericValue 1.500000e-1
    ]
  ] ;
  ssn:observationResultTime [
    a time:instant ;
    xsd:dateTime "2016-9-13T21:32:52"
  ] ;
  ssn:observerBy
<http://iotservice.hueuni.edu.vn/station/DN1> .
[] a qudt:Density .
```

Fig. 6. Sample RDF described Air Polution : PM10 *(http://iotservice.hueuni.edu.vn/data/DN1/PM10_1476893293)*

Data stored in a Graph Store server and can be access through the service. Output data can be access in normal web service data like JSON, JSON-LD, XML, HTML for other formal web application. For example, a map application use IoT service data combined with Google API v3. We use mashup technology that combine our service data with third party service like google map, open location, easy graph to build a sample high level application.

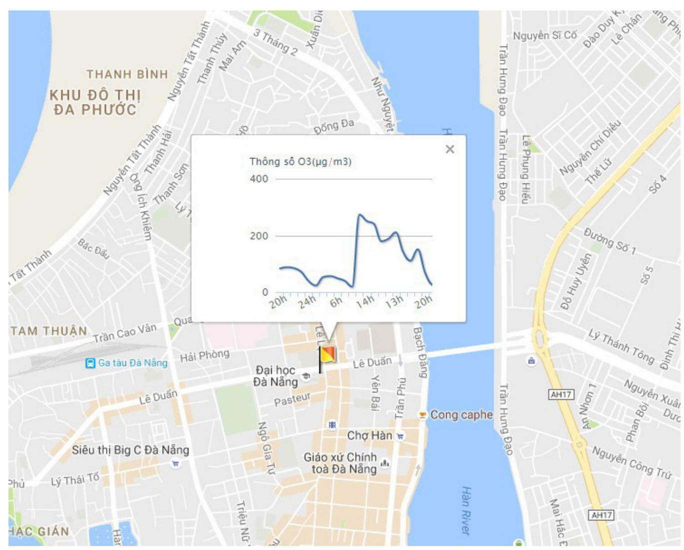


*Fig. 7.* Da Nang Air Quality Sample *(http://iotservice.hueuni.edu.vn/data/map)*

## VI. Discussions of Preliminary results

By using a light-weight data model using Semantic Web technology, sensors data not only represent and ready for reuse but also provide a reasoning base and knowledge extract from rich semantic data for other services and IoT applications. Processing semantic streaming data over Linked data [5] is a promising field for Semantic web of Things applications.

We have presented a set of URIs to identify sensors and sensor data. Our output of this service is multi data model that enable for service and application. Raw sensor data is ready for other application and semantic query, reasoning. This service support multi rich data type output: JSON, JOSN-LD, GRAPH, RDF, XML. For evaluate, we implemented the real dataset from fixed automatic stations in Da Nang, Ha Noi, Nha Trang. Our framework real-time listening and initiative get data from the station when it has request from clients. The experiment is run on o server in Hue University Data Center (http://iotservice.hueuni.edu.vn), this server address is also our IoT Service system. For the Graph Store server, we design a dedicated UNIX server and not public address for security issue, this sever is connected with IoT server and receive graph store and send graph on frontend server demands.

## Outlook and Conclusions

We have presented a light-weight proposal for presenting environment data by using a semantic middleware based on SSN ontology for environment events management and query. Our objective is hide the heterogeneity of data sensor by using a common model easy for query, reasoning data. This architecture is proposed to demonstration how data transform from raw to rich semantic description. From raw data, all sensor metadata and their measurement can be published, indexed, queried and most important this data can be discovered, reasoning and reused for other applications and services. The novel of our proposal is reuse the knowledge and design a lightweight model for sensor data, this model can combine data measurement and other sensor metadata like Geo location, Station, Sensor information.

Our plan is building a Semantic IoT framework for processing data and provide a IoT common schema for sensors data. Our framework attention will be given the light-weight data models and data mapping templates can help solving semantic heterogeneity in IoT sensors and the last point is the IoT data quality. The future work will focus on the data template generation for heterogeneous source that data can transform automatically with defined cross-domain knowledge.

## References


[1] A. Eisfeld and D. A. Mcmeekin, "Complex Environment Evolution," pp. 1–6, 2013.

[2] W. Wang, P. Barnaghi, G. Cassar, F. Ganz, and P. Navaratnam, "Semantic sensor service networks," *Proc. IEEE Sensors*, pp. 3–6, 2012.

[3] B. Thuraisingham and K. W. Hamlen, "Secure Semantic Sensor Web and Pervasive Computing," *Sens. Networks, Ubiquitous, Trust. Comput. (SUTC), 2010 IEEE Int. Conf.*, pp. 5–10, 2010.

[4] M. Botts, "Sensor Model Language (SensorML) v2.0," *OGC Implementation Specification*, 2013. [Online]. Available: http://www.opengeospatial.org/standards/sensorml.

[5] M. Compton, P. Barnaghi, and L. Bermudez, "The SSN Ontology of the Semantic Sensor Networks Incubator Group," *J. Web* pp. 1–6, 2011.

[6] D. B. Manu Sporny, D. B. Dave Longley, K. A. Gregg Kellogg, G. U. of T. Markus Lanthaler, and Niklas Lindström, "JSON-LD 1.0," 2014. [Online]. Available: https://www.w3.org/TR/json-ld/.

[7] M. Lanthaler and C. Gütl, "On using JSON-LD to create evolvable RESTful services," *Third Int. Work. RESTful Des.*, no. April, pp. 25–32, 2012.

[8] J. F. Sequeda and O. Corcho, "Linked Stream Data : A Position Paper," pp. 148–157, 2009.

[9] E. Bash, "R2RML: RDB to RDF Mapping Language," *PhD Proposal*, 2015. [Online]. Available: https://www.w3.org/TR/r2rml/.

[10] D. Tomaszuk, "Managing a graph store in the Linked Data and REST style," *Media*, no. October, pp. 1–5, 2016.

[11] R. G. Raskin and M. J. Pan, "Knowledge representation in the semantic web for Earth and environmental terminology (SWEET)," *Comput. Geosci.*, vol. 31, no. 9, pp. 1119–1125, 2005.