Unleashing Sensor Data on Linked Open Data - The Story So Far

Shah Khusro, Shaukat Ali, Azhar Rauf, Saeed Mahfooz, Syed Rahman Mashwani, Akif Khan, Fouzia Jabeen

Department of Computer Science, University of Peshawar, Peshawar, Pakistan
{khusro, shoonikhan, azhar.rauf, saeedmahfooz, syed.rahman, akif, fouzia.jabeenphd}@upesh.edu.pk

Abstract: The first decade of the 21st century has seen tremendous technological developments in sensor networks and increase in the potentialities of creating energy efficient and low cost sensor embedded networked devices. These significant improvements have resulted in the deployment of a very large number of sensing devices in various domains for capturing a huge amount of data. To make the captured data potentially useful for applications, it needs to be openly accessible to applications in an understandable pattern and linked with other related open data sources instead of being locked inside organizations. However, the data produced by sensors normally is in different formats and lacking of semantic to describe their meanings. This poses significant problems in accessing and using sensor data in applications, and linking with other related data sources. To solve these problems, sensor data needs to be annotated using Semantic Web technologies and published over the LOD cloud using the Linked Data principles. The resulting integration opens novel ways to both industry and academia for designing useful web services and applications. The efforts in this direction are not well organized and well coordinated and make it difficult for a researcher to know state of the art and precisely identify an issue and follow a clear course of investigation. This paper covers the gap by presenting a comprehensive overview of the research efforts committed by the relevant research communities for annotating, uploading, and linking sensor data with data sources on the LOD using the Semantic Web technologies. It also identifies limitations of the approaches followed and concludes with a list of recommendations for future research.

[Shah Khusro, Shaukat Ali, Azhar Rauf, Saeed Mahfooz, Syed Rehman, Akif Khan, Fouzia Jabeen. **Unleashing Sensor Data on Linked Open Data - The Story So Far.** *Life Sci J* 2013;10(4):1766-1786]. (ISSN:1097-8135). http://www.lifesciencesite.com. 232

Keywords: Linked Open Data; Semantic Web Enablement; Web Ontology Language; metadata; Resource Description Framework; Sensor Web; Semantic Sensor Web

1. Introduction

Sensors have now become ubiquitous in nature, by envisioning and deploying a multitude of sensors all over the planet for capturing information about a number of real world phenomena (Phuoc and Hauswirth, 2009). Moreover, sensors are now common to be found in human's daily usage devices and environment. For example, state-of-the-art smartphones (i.e. comes with several built-in sensors such as GPS, WLAN, accelerometer, digital compass, camera, microphone etc), vehicles, home appliances (i.e. microwave oven etc), and standalone computer systems are equipped with multitude of sensing devices which could be effectively used in different domains. Deploying sensors in environment at large scale and using them as data sources to produce real-time data emerges novel sources of information which could be used fruitfully in a number of disciplines including civic planning (i.e. traffic management etc), meteorology (i.e. weather forecasting etc), medical sciences (i.e. patient care etc), wildfire detection, ambient intelligence gathering, satellite imaging (i.e. monitoring earth and space observations), and supply chain management (Phuoc and Hauswirth, 2009)(Sheth et al., 2008). Such a large number of sensors will produce tens and tens of data in real-time, which will be either in the raw format or in the format which could not be used by the applications directly. To enable the data generated by sensors suitable for decision making, some further processing would be essential. According to Gartner: *"By 2015, wirelessly networked sensors in everything we own will form a new Web. But it will only be of value if the `terabyte torrent' of data it generates can be collected, analyzed and interpreted." (Raskino et al., 2005).*

A sensor network is a collection of sensor which would be spatially distributed and accessible through computers, and could be used to monitor environmental conditions (e.g., humidity, temperature, vibration, noise, pressure, pollution, rain, and movement etc) at diverse geographic locations. A Sensor Web is a collection of sensor networks accessible through Web. Real-world application can discover and access real-time as well as archived sensor data from the Sensor Web using pre-defined specialized application programming interfaces (APIs) and networking protocols (Botts et al., 2007).

Organization normally builds their sensors infrastructure to collect data of their interests, and lock them inside their organizational boundaries for using in their specific applications (Phuoc and Hauswirth, 2009). To potentially exploit sensors generated data as a novel and prime resource of information will demand its integration into the existing Web information space, which by itself is a difficult and laborious task. But, this integration of physical world information with the existing resources and services on the Web will not only revolutionaries the market trends of constructing novel type of front-end applications but will also changes the attitude of organizations and peoples of using Internet services and applications in their day-to-day lives (Barnaghi et al., 2010). Due to the lack of inter sensor networks communication, the tasks of accessing the sensors data sources and integration of vast amount of data generated by sensors are pretty much cumbersome in their own capacity (Phuoc and Hauswirth, 2009). The recent decrease in the price of the commodity sensors will not only encourages for the deployment of numerous sensor networks but will also demand for a platform where to publish and share sensors data with reduce level of cost, less complexity of sensors data integration, and easy to access the integrated sensors data.

Several research projects have contributed their interests to envisioning and integrating large-scale sensor networks into the Web to reduce or solve the problem of integrating published sensors data online such as EarthScope [1], SensorMap [2], EarthCam [3], SENSEI [4], SensorWeb [5] (Phuoc and Hauswirth, 2009)(Barnaghi et al., 2010). But, they have several problems: (1) integration of these data sources and their accessibility to applications is difficult, (2) providers does not or rarely supports sensors discovery in large-scale sensor networks, (3) users with no or low-level programming skill faces problems in the integration of sensors data into their applications. To solve these problems, activities from Open Geospatial Consortium (OGS) [6] as well as Sensor Web Enablement (SWE) have been contributed to the creation of service layer and data structures for sensors and actuators networks (Botts et al., 2007). Among the other, these are some of the enduring efforts to encourage applications development for the large-scale sensor networks. Collaboration, semantic interoperability, and scalability should be the key features in scheming a large-scale sensor network for competent resources distribution and information exchange. They will provide potential for creating network resources with the abilities to collect data and services from the physical world and Web. Interlinking sensors data (representing a physical world phenomenon) with the web will contribute into fulfilling the vision of Semantic Web of creation of networked knowledge infrastructure (Barnaghi et al., 2010).

Sensors generated data is in raw and diverse formats (e.g. JSON. RSS etc), which is rarely associated with metadata to determine its meaning. Sensors generated data meanings includes features of interests, measuring device’s specifications, measuring conditions, accuracy, location, and scenario of measurements etc (Phuoc and Hauswirth, 2009)(Le Phuoc).Metadata is essential for effective management of sensors data (Sheth et al., 2008). Metadata will help users who have to confront with multitude of sensors and tones of sensors data, particularly in situations when a user is not sure about his search, and starts his search with more general or relevant concepts and using the semantic descriptions and their relations, makes his search narrower. Using of Semantic Web technologies in Sensor Web is helpful for several reasons (Keßler and Janowicz, 2010): (1) enriching sensors data by semantic annotation, (2) using onotlogies to make provided data unambiguous and enrich with machine readable descriptions. Spatial, temporal, and thematic information required for determining and evaluating sensors data can be easily provided by a semantically rich sensor network (Le Phuoc). Semantically annotated sensors data can be understood by applications coherently, consistently, and accurately.

The Sensor Web Community has recently dramatized the Linked Data rules (Bizer et al., 2009): using URIs for reference, using HTTP URIs for look up, using RDF for storage, using SPARQL for access as well as querying, and providing links using URIs. The Linked Open Data uses Semantic Web technologies to provide an infrastructure to publish data regarding any domain and interconnect them through defining relations between ontologies or schemas with other existing resources on the Web. Publishing sensor data on LOD will ensure publically accessibility to sensors data along with their related metadata. This will not only facilitate in finding other relevant data/information but would also provide ease in interconnecting and integrating sensors data with data from diverse sources and communities. This access to vast amount of interconnected information will provide an excellent platform for the development of useful applications and services for the Web of the future.

This paper presents a comprehensive overview of the research committed by the organization, academia, and research communities for annotating, publishing, interlinking sensor generated datasets with datasets on the Linked Open Data using Semantic Web Technologies. To demonstrate, we have classified this paper into a number of sections. Section 2 gives a detail overview of Linked Open Data and its state-of-the-art. Section 3 describes a detail list of motivational

---

[1] http://www.earthscope.org/
[2] http://atom.research.microsoft.com/sensewebv3/sensormap/
[3] http://www.earthcam.com/mapsearch/
[4] http://www.ict-sensei.org
[5] http://research.microsoft.com/en-us/projects/senseweb/
[6] http://www.opengeospatial.org

scenarios. Section 4 presents a detail discussion of research conducted for linking sensors data to the Web of data. In Section 5 the existing research work in the area of linking sensor data to Linked Open Data is elaborated. In Section 6 the existing application systems developed by researchers utilizing linked sensor data is discussed. Section 7 suggests some recommendations and concludes the paper. Key contributions of this paper includes:

- The key contribution is the study, collection, summarization, and presentation of all of the on hand research literature encompassing Linked Open Data and its state-of-the-art, sensors data semantic annotation, publishing sensors data on Linked Open Data, sensors data applications. Apart from these, we have also highlighted the on hand research efforts contributed in these fields by the research communities. We have described the functionalities and working of each system which make use of the sensors data and referenced them in the paper.
- The novel contribution is the categorization of the on hand research work contributed to sensors data semantic annotation, and categorization of sensors datasets available so far. Annotating sensor data semantically is divided into three broad categories: (1) using ontologies, (2) using LOD, and (3) using both of them. Sensor datasets are divided into sensors oriented and sensors related categories.
- A comprehensive list of motivational scenarios is described for helping researchers in understanding the importance of sensors data implications in the design of useful applications and systems in the future.
- The knowledge is organized and presented in a way to boost up the interest of new researchers in the area while providing them necessary initial knowledge to understand and add contribution to the area.
- A number of recommendations are highlighted to serve as new dimensions for the researchers in this domain.

2. Linked Open Data (LOD)

A great deal of information describing numerous domains could be made accessible by using Semantic Web technologies for publishing data as well as interlinking them with other resources on the Web. Tim Berners Lee coined the idea of Linked Data in 2006 and considered it as transformation of Web of Documents into Web of Data[7]. Technically, Linked Data refers to the publishing of structured data on the Web in such a way that it is formal (machine readable), explicit (meaning is clearly defined), and can be linked to and linked from external datasets (Bizer et al., 2009). Berners Lee identified the basic guideline for building a Linked Data structure including[8]: (1) URIs should be used for things identification, (2) HTTP URIs should be used to provide HTTP access to the URIs for look-up, (3) Using well defined standards (such as RDF, OWL, and SPARQL etc) for data storage, accessing, and querying through URIs, and (4) to get more complete list of knowledge, data should be interlinked using URIs (Bizer et al., 2009).

Linked Data uses two fundamental Web technologies: URIs (Uniform Resource Identifier) and HTTP (Hypertext Transfer Protocol) (Bizer et al., 2009). To accomplish the idea of Web of Data, instead of using HTML to link pages, resources on the Web should be annotated and interconnected via links in a way that can be queried and interpreted by discovery and search agents. Users can uses Linked Data to start navigation from one data source and continue to browse a vast number of resources from other data sources by following the machine readable links (e.g. RDF links) connecting them (Barnaghi et al., 2010).

W3C Semantic Web Education and Outreach Working Group has initiated Linked Open Data community project with the objective of using Semantic Web technologies (e.g. RDF, RDFs, and OWL etc) (Breitman et al., 2006) to support the Web of Data. RDF is a generic, graph based data model showing how to structure and link resources to describe things in the real world. RDF links takes the form of RDF Triples: subject, predicate, and object. The subject part of a RDF triple would be URI reference corresponding to the namespace of one dataset, whereas the corresponding object would be either a URI reference corresponding to either the namespace of the same dataset or another dataset, or would be a string literal respectively. The property (predicate) indicates the relationship between the corresponding subject and object and would be also represented by a URI in the namespace of a dataset. The RDF Vocabulary Description Language (RDFs) and the Web ontology Language (OWL) provides constructs for defining vocabularies to describe the real-world entities and their links with each other in a more detailed and meaningful fashion. A vocabulary contains classes and properties and is represented in RDF using the concepts and constructs from RDFs and OWL (Bizer et al., 2009). Using HTTP URIs to identify resources, HTTP protocol for dereferencing URIs, and RDF as data model for describing resources, the Web of Data (Linked Data) can be directly build using the general architecture of the Web.

The W3C SWEO Linking Open Data community[9] project was initiated in 2007 and by

---

[7] http://www.w3.org/DesignIssues/LinkedData.html

[8] http://www.w3.org/DesignIssues/LinkedData.html

[9] http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project_Description

September 2011 reports, the project succeeded in publishing and interlinking 295 datasets. These dataset have over 31 billion RDF triples, interlinked by over 504 million RDF links. The project includes several open datasets accessible on the Web, for example Wikipedia[10], Wikibooks[11], GeoNames[12], MusicBrainz[13], WordNet[14], DBLP Bibliography[15], and several others which are published under the Creative Commons[16] and Talis[17] licenses. The project is aimed to extend the Web by publishing a range of datasets as RDF, establishing RDF links among data items from diverse datasets, and making them accessible through query interface (SPARQL endpoints). With growing tendency, public and government organizations have published their data as liked-data recently. For instance, UK government has recently provided linked-data, whereas, the published datasets are accessible through SPARQL endpoints[18] (Barnaghi et al., 2010).

2.1. Linked Open Data Cloud State-of-the-Art

In the past few years, a number of open datasets have been build and published using Linked Data format by the Linked Open Data community projects, individuals, and organizations. Initially research projects and Web enthusiasts showed interest in the Linked Data best practices[19]. These third-parties developed a number of Linked Data wrappers for the existing Web APIs, converted existing datasets into RDF, and published over the Web. With increase in the Linked Data technologies, data producers started developing and providing access to their datasets. By August 2011, producer by themselves published 38.57% (113 out of 295) of datasets in the LOD cloud, while the third-parties published the rest of 61.43% (180 out of 295) of datasets. At present, Linked Data technologies are potentially utilized for sharing data describing a wide range of diverse domains. Table 1 show the quantity of triples and the quantity of RDF links per domain.

Web of Linked Data envisioned helping applications in discovering and integrating required data from the global Web of interconnected data sources. To bring this vision into reality, data providers are required to publish their data according to the set of best practices of Linked Data. A data source fulfilling these best practices would be included in the LOD cloud and almost all of the data sources in state-of-the-art LOD cloud have passed this pre-condition. A data source in the LOD cloud is accessible if it is connected with other data sources through outgoing RDF links or being the target of outgoing RDF links from the other data sources in the LOD cloud. Table 2 show the categorization of LOD cloud datasets using absolute number of outgoing RDF links. Table 3 show the categorization of LOD cloud datasets which are target of the outgoing RDF links of the other datasets.

Data providers most often uses the terms defined in the widely deployed and verified vocabularies for representing data in the Linked Data as well as making them convenient for applications to understand. W3C has invested its great contribution in the development of standardized vocabularies and has been succeeded in achieving landmarks by developing a number of vocabularies including RDF, RDF Schema, and OWL, whose terms are nearly used by all of the datasets available and linked in the current LOD cloud. Figure 1 depicts the division of the profoundly used vocabularies.

Table 1. Quantity of triples and quantity of RDF links per domain[19]

Domain Name	No. of Datasets	No. of Triples	Triples Percentage	No. RDF Links	RDF Links Percentage
Media	25	1,841,852,061	5.82	50,440,705	10.01
Geographic	31	6,145,532,484	19.43	35,812,328	7.11
Government	49	13,315,009,400	42.09	19,343,519	3.84
Publications	87	2,950,720,693	9.33	139,925,218	27.76
Cross-domain	41	4,184,635,715	13.23	63,183,065	12.54
Life sciences	41	3,036,336,004	9.60	191,844,090	38.06
User-generated content	20	134,127,413	0.42	3,449,143	0.68
Total	295	31,634,213,770	100	503,998,829	100

Table 2. LOD datasets categorization using the absolute number of outgoing RDF links[19]

Outgoing Links		Number of Datasets	Percentage
From	To		
Upto 0		30	10.17 %
1	1,000	90	30.51 %
1,000	10,000	58	19.66 %
10,000	100,000	45	15.25 %
100,000	1,000,000	43	14.58 %
More than 1,000,000		29	9.83 %

---

[10] http://www.wikipedia.org/
[11] http://www.wikipedia.org/
[12] http://www.geonames.org/
[13] http://musicbrainz.org/
[14] http://wordnet.princeton.edu/online/
[15] http://www.informatik.uni-trier.de/~ley/db/
[16] http://creativecommons.org/
[17] http://tdnarchive.capita-libraries.co.uk/tcl
[18] http://data.gov.uk/sparql
[19] http://www4.wiwiss.fu-berlin.de/lodcloud/state/

Table 3. LOD datasets categorization using the number of data sources target of outgoing RDF links[19]

Number of Linked Datasets	Number of Datasets	Percentage
> 10	27	9.15
>= 6 AND < 10	17	5.76
= 5	5	1.69
= 4	19	6.44
= 3	38	12.88
= 2	62	21.02
= 1	98	33.22

The URIs used to represent resources in the data published on the Web of Linked Data should be de-referenceable into RDF description. Two methods have been provided for accessing the data from the LOD cloud. Firstly, SPARQL endpoints, which provide support and potential to the users of executing expressive queries against the datasets for data retrieval and checking consistency. Second, RDF dumps, which is a complete RDF dataset downloadable from a disconnected URL. Among all of the 295 data sources, 68.14% (201data sources) provides SPARQL endpoints and 39.66% (117 data sources) provides RDF dumps. More detail information about LOD cloud is available at[20].

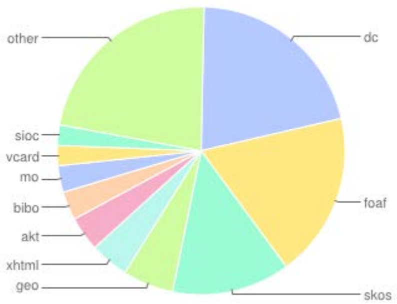


Figure 1. Distribution of most widely used vocabularies[21]

3. Motivational Scenarios

In the recent years a number of weather organizations collected a vast amount of data about major catastrophic situations such as Hurricane Ike (2008) and North American blizzard (2010) from the sensors deployed around the world, especially in the United States. The provision of globally open availability of the collected data can be helpful in predicting and stopping several catastrophic situations which could be destructive otherwise. But, retrieval of sensors collected data from traditional storage frameworks could be a trivial task due to a number of reasons such as heterogeneity of underlying architectures, and applications etc.

Semantic Sensor Web highlights annotating sensors data using semantic metadata annotation techniques to furnish contextual information necessary for situational awareness. Semantic metadata annotation is taken to be helpful in determining suffices to complex queries encompassing spatial, temporal, and geographical domains. Linked Open Data cloud on the Web, due to its unique nature is a global decentralized information space, allowing applications to access and use available data in interconnected datasets for solving plenty of real-world problems. With the availability of sensor data, as datasets, will open a new paradigm of applications development requiring discovery of corresponding datasets and usage of data contained. A number of scenarios has been posted by (Phuoc and Hauswirth, 2009)(Leggieri et al., 2011)(Le-Phuoc et al., 2010)(Patni et al., 2010) concentrating on the usage of global sensors datasets for solving real-world problems.

- Sensor data can be annotated with contextual information for linking with Linked Open Data (LOD) cloud. A driver can drive faster if he follows the routes suggested by his particular GPS car navigation system. The route suggestions will be based on composite set of data accessed from LOD's datasets such as information about the hilly surrounding area (from Geography LOD datasets), information about nearby road works (from Government LOD datasets), and information about ongoing social events in the locale (from Media LOD datasets).
- A farmer normally browses national weather agencies websites to get information that might be of potential danger for his seeds, which is a difficult and laborious task and can sometime result in destructive situations because of not accessing the important information in real-time. By linking sensors data from such agencies to the LOD, automatic notifications applications can be build which will take into account the relevant sensors data (e.g. social media, weather stations, and streams etc) and upon meeting the certain conditions (e.g. temperature, humidity, and wind force etc) throws an automatic text message to farmers, informing them about potential dangers in advance.
- In cataclysmic situations (i.e. epidemic outbreaks etc), the officials mostly uses the information posted by the people on social media such as Twitter, and Facebook etc to take appropriate actions such as identifying the scope of infection, understanding the nature and reason of outbreak, evacuating people, and providing of required medicines etc. By combining the users generated contents from the

---

[20] http://www4.wiwiss.fu-berlin.de/lodcloud/state/

[21] http://www4.wiwiss.fu-berlin.de/lodcloud/state/

social media with the appropriate sensors data (already deployed in the area) from the LOD can further help the officials in preventing and countering the disease.

- Using Semantic Web technologies sensors datasets can be linked with LOD, which can be queried using query languages (e.g. SPARQL etc) to extract a complete set of real-time information about a phenomenon. For example, to get information about a blizzard we can issue a query encompassing the blizzard time duration, the blizzard location, and the sensors deployed in the area during the time period. The retrieved information can be used by applications such as analyzing the information to find out the exact reason of blizzard etc.
- A user's presence can be managed by integrating his physical presence information from different relevant sensors data (e.g. Geographic Position Sensor, Wireless LAN, Bluetooth, RFID, activity detection sensors, and noise sensors etc) from LOD with his virtual presence information from software such as calendar information, online status in Skype, and collaborative environments etc. For example, combining the information from user's calendar with his physical location sensor data, all of his appointments can be automatically changed if the system infers that all of the participants could not reach the meeting location in time. Upon determining a user physical location or activity (e.g. being in a meeting etc) from several sensors (e.g. audio sensors or RFID readers in the user's office etc), his online status on Skype can be changed automatically.

In addition to these, several other straight-forward and complex scenarios can be demonstrated, if the sensors data is flexibly and transparently integrated with other sources of information. Linked Data approach can be used to efficiently link sensors data with LOD and to make them globally accessible ultimately.

4. Linking Sensor Data to the Web of Data

Recent technological successes in the development of small size, low-cast, and energy efficient sensing devices has lead to the potential of building large-scale sensor networks capable of observing, capturing, and measuring a number of physical phenomenon. The captured information can be processed for making them suitable to be used in varied services and applications belonging to different areas including health care, Geographic Information Systems (GISs), and smart homes etc.

In the last few years, a significant amount of the research community have contributed their efforts to the development of large-scale sensor networks (e.g. SENSEI project [22], and SensorWeb [23] etc), and developing industrial standards for sensors data description (e.g. Sensor Model Language standard proposed by the OGC's Semantic Web Enablement activity). XML is used as the main tool for representing sensors data description, which is not expressive enough to support semantic interoperability and relating the described resources to the existing Web of Data (Wei and Barnaghi, 2009). To leverage the data collected from heterogeneous sensors needs to be integrated and made available to various services and applications, Semantic Sensor Web (Sheth et al., 2008) combines the technologies of Sensor Web with the technologies of Semantic Web. SSW extends the specifications of OGC and SWE with the Semantic Web technologies for providing more enhanced spatial, temporal, and thematic semantic annotations to sensors description for facilitating access to sensors data. SSW relies purely on Semantic Web technologies such as Web Ontology Language (OWL), and Semantic Web Rule Language (SWRL), which adds increased levels of un-required complexities and creates new problems instead of solving them (Keßler and Janowicz, 2010). Researchers (Keßler and Janowicz, 2010)(Patni et al., 2010)(Balazinska et al., 2007) have proposed a more light-weight approach based on Linked Data, to represent sensors data in RDF which is not only enough to expose sensors data to a wide community of users and applications but allows its integration with other resources in the Linked Open Data cloud. Recent implementation ranges from the static conversion of datasets to the automatic on-the-fly conversion from OGC services to RDF format using software tools (Patni et al., 2010).

4.1. OGC's Sensor Web Enablement (SWE)

OGC (Open Geospatial Consortium) (Botts et al., 2007) is a worldwide conglomerate of academic, industry, and government organizations who contributes to the development of open standards for location and geospatial services collaboratively[24]. OGC standards activities focus on the building of Sensor Web Enablement (SWE) framework which encompasses sensors, sensors networks, and Sensors Web. SWE framework is a group of open standards which covers sensors of all types including Web connected such as air pollution monitors, flood gauges, webcams, mobile heart monitors, stress gauges on bridges, satellite-born earth imaging devices, and many other sensors and sensors systems.

SWE architecture defines a number of models, encodings, and services to promote the realization of interoperable and scalable service-oriented networks

---

[22] http://www.ict-sensei.org/

[23] http://research.microsoft.com/en-us/projects/senseweb/

[24] http://research.microsoft.com/en-us/projects/senseweb/

consisting of diverse sensor systems and user applications. The OGS's SWE initiative develops standards to add additional functionalities in Sensor Web such as discovery of sensor systems, observations, and observations processing which stratifies an application's or user's distinct needs; estimation of a sensor's capabilities and quality of measurements; access to sensors parameters to automatically enable software to process and geo-locate observations; retrieval of real-time sensors observations and converting into standard encodings; querying sensors for acquiring observations of interests; publishing and subscription to alerts to be issued by sensors and sensor systems based upon certain criteria. In addition, SWE initiatives describe several encoding for sensors observations and several standard interface definitions for accessing SWE's Sensor Web functionalities through web services.

OGC standard have defined seven OpenGIS specifications in Sensor Web Enablement framework:

- **Observations and Measurements Schemas (O&M)[25]:** Instead of providing support for varied data formats defined specifically for sensors and communities, O&M standard defines a theoretical model along with a XML schema for representing and exchanging measured and observed results as well as for characteristics used for sampling while attaining observations. O&M standard make use of flexibilities and extensibilities inherently offered by XML for effectively packaging of large amount of data as ASCII or binary block. This standard defines eight different units of measuring observations. Recently, OGC's Observations and Measurements v2.0 has been promulgated as ISO/DIS 19156 standard.
- **Sensor Model Language (SensorML)[26]:** SensorML[27] defines information model, XML schema, and encodings for discovery and tasking of sensors accessible through Web, and for describing any process (exploitation of observations and measurements of any sensor system plus any after measurement processing). Instead of describing hardware details, SensorML describe the functional details of a sensor system. SensorML considers both sensors systems and a system's elements such as sensors, actuators, detectors, filters, platforms, and operators etc as processes and each process model has inputs, outputs, parameters, and methods along with large amount of metadata helpful for human support, detection, system constraints identification, supplying contacts, and identifying taskable requirements.
- **Transducer Markup Language (TransducerML)[28]:** Sensors and actuators are the subsets of a transducer. TML standard defines a protocol for application and presentation layer which uses XML to define methods, encodings, and message formats for representing information related to sensors and sensor systems, and interchanging as well as capturing of live streaming, archived data, and future data produced or captured by a sensor system. A transducer system can have different types of numerous transducers, receivers, actuators, transmitters, and procedures. TML provides accurate and efficient approach to transport, capture, and archive transducers data without requiring a client requiring having any prior knowledge of TML enabled system. TML protocol completely defines both the transducers data and the transducers systems itself. TML standard is scalable, reliable, explicit, and can be used in any sensor system consisting of any number and types of sensors and actuators.
- **Sensor Observation Service (SOS)[29]:** SOS standard describes a consistent Web service interface which supports querying any sensor or sensor system (such as remote, in-situ, mobile, and fixed etc) for observations, sensors metadata, and representation of observed features. SOS standard delineate an Application Programming Interface (API) for registering new sensors, removing existing sensors, operations for new sensors observations, retrieving sensors observations and measurements, and two types of bindings: KVP and SOAP. SOS serves as an intermediate among clients and observations repositories, and can be used by clients (after registering) to capture metadata information describing related sensors, platforms, methods, and metadata related to sensors observations.
- **Sensor Planning Service (SPS)[30]:** SPS standard provides an efficient interface for supporting queries related to retrieving information about capabilities of associated sensors and how the sensors can be tasked. More specifically this interface describe the capabilities of SOS to support queries related to: (1) finding the viability of a transducer planning request; (2) request submission, committing, and roll backing; (3) request status determination and processing such as updating, and cancelling etc; (4) and determining data related to other OGC Web services providing access to the information accumulated by the requested job.

---

[25] http://www.opengeospatial.org/standards/om

[26] http://www.opengeospatial.org/standards/sensorml

[27] OGC adopted and extended SensorML from NASA and CEOS projects.

[28] http://www.opengeospatial.org/standards/tml

[29] http://www.opengeospatial.org/standards/sos

[30] http://www.opengeospatial.org/standards/sps

- **Sensor Alert Service (SAS):** SAS standard specifies alert as a special kind of notification indicating the occurrence of an event at an object of interest. This standard define an interface for retrieving information about the capabilities of Sensor Alert Service, for identifying nature of offered alerts, the protocols used, and provides facilities for subscription to specific alert types. SAS need users to be registered to receive alerts related to requested events such as weather events or earthquakes etc.
- **Web Notification Service (WNS):** WNS standard defines an open interface allowing users to conduct asynchronous communication with one or more other services. WNS supports two types of communications. First, "one-way communication" forwards information to users without requiring a response. Second, "two-way communication" forwards information to users requiring some type of asynchronous response.

In spite of OGC's Sensor Web Enablement (Botts et al., 2007) project significant efforts for developing a set of standard languages and Web services interfaces for managing Web accessible sensor data, it still have certain shortcomings. All of the SWE standard languages are XML-based which effectively provides syntax-level interoperability but does not support the semantic-level interoperability which is significantly needed for advanced integration and analysis. In order to solve this challenge, research community has recently made many attempts of combining Sensor Web and Semantic Web technologies and laid the foundation of Semantic Sensor Web (Sheth et al., 2008).

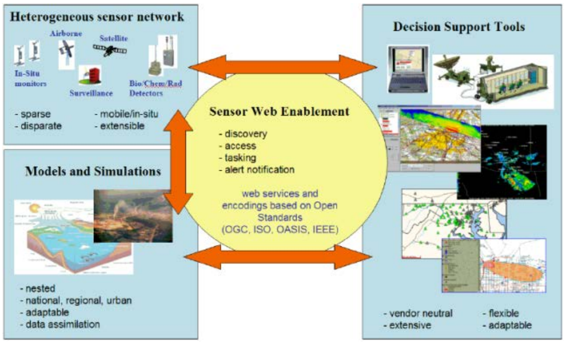


Figure 2. Role of the Sensor Web Enablement (Botts et al., 2007)

4.2. Semantic Sensor Web

Currently millions of sensors of different types with varied capabilities (e.g. mobility, range, and maneuverability etc) are collecting avalanches of data related to our environment. Today's networks can accommodate sensor of different types but lack of these networks integrations and communication can create a catastrophe of losing these avalanches which could be used productively otherwise. To alleviate this problem, sensors data needs to be associated with semantically enriched spatial, temporal, and thematic metadata for defining contextual information required for situational awareness. Associating sensors with semantic metadata provides guidelines for Semantic Sensor Web (SSW)[31] (Sheth et al., 2008). Understanding the importance of SSW, World Wide Web Consortium (W3C) started Semantic Sensor Network Incubator Group (SSN-XG) with the objective *"to begin the formal process of producing ontologies that define the capabilities of sensors and sensor networks and to develop semantic annotations of a key language used by services based sensor networks"* [32].

SSW leverages and extends the standardization efforts of OGC and World Wide Web Consortium (W3C) by incorporating Semantic Web technologies for providing improved description and accessibility to sensors information. It was believed that Semantic Sensor Web will not only improve interoperability will also introduce a new dimension of applications development, services, and functionalities with data from heterogeneous sensors including situational awareness. Associating sensors data with spatial, temporal, and thematic metadata by semantically rich sensor networks can make management, discovery, and analyzing of sensor data pretty much easy.

Semantic Sensor Web enhances the SWE's existing standard sensor languages through adding semantic annotation to sensors data. In other words, SSW is an OGC-type Sensor Web enriched with semantic annotation, inferencing, and querying (Compton et al.). Semantic annotation will not only add meaningful descriptions to sensors data will also improve accessing them as compared to SWE alone. Semantic annotation will serve as a linking mechanism to fill the gap between the SWE's basic synthetic XML-based metadata standards and Semantic Web's RDF/OWL-based metadata standards. SSW uses entirely Semantic Web technologies (i.e. ontologies and rule-based searching etc) to improve interoperability, analysis, and reasoning over sensors data represented in varied formats.

SSW has potential applications in several areas such as biometrics, weather forecasting, oceanography, EventWeb, and videos on the Web. Complex queries using simple weather readings (i.e. requesting data related to freezing or blizzard situations at a particular place and time etc) can now be fluently executed by leveraging SSW Semantic annotations. A

---

[31] http://knoesis.wright.edu/research/semsci/application_domain/sem_sensor/

[32] http://www.w3.org/2005/Incubator/ssn/

typical application is the YouTube video originated from Ohio State Petrol in-dash cameras containing temporal data, encoded in SensorML, and semantically annotated using OWL-Time ontology. Videos can be retrieved through queries using semantic temporal concepts such as contains, within, and overlap, which can be positioned on Google Map and played within an information window.

5. Linked Sensor Data

Linked Open Data vision behind linking data is to increase the usefulness of data by interlinking it with a large amount of other openly available and accessible data (Wei and Barnaghi, 2009). The prime object behind publishing sensors observations and measurements as Linked Data is to ensure their availability beyond the spatial data infrastructure, accessibility using HTTP-based URIs, and re-usability of sensors data through supporting their integration and fusing (Corcho and García-Castro, 2010). Materialization of sensors data as Linked Data enables sensors data providers to publish and connect the sensors descriptions to everlasting vast collection of data already existing on the Web using the principles of Linked Data. With publishing of sensors data and relating it with other resources on the Web using sensors data features, applications can integrate information from different communities and sources for a number of reasons including: drawing conclusions, enable smart environments, create business intelligence, and support automated decision making systems etc. Linked sensors data can be queried, retrieved, analyzed, reasoned, and inferred like others, therefore, creates a novel open platform for publishing and consuming sensors data in an interoperable manner (Barnaghi et al., 2010).

5.1. Sensors Data Annotation

Providing universal descriptions to sensors data greatly improves the applications developers' throughput without being involved in dealing with the complex issues of technological heterogeneity in diverse platforms (Wei and Barnaghi, 2009). Describing sensors data using metadata and semantic annotation from scalable and heterogeneous environments will empower diverse communities to utilize and exchange the emergent information in a collaborative environment. In the past few years, Semantic Web community researchers have established standards (i.e. formal knowledge representation frameworks, and ontologies etc), which have the potential of modeling various sensors data at several levels to affirm interoperability. Moreover, sensors data can be integrated with rich description in the form of semantic annotation, which facilitates advance query and retrieval tasks by performing logical reasoning using the semantically enriched sensors data.

To facilitate advance query and reasoning, Sheth et al. (Sheth et al., 2008) augmented the vision of Semantic Sensor Web, advocating the addition of semantic annotations to sensors data in terms of spatial, temporal, and thematic scope.

- **Spatial Information:** Sensors data can be associated with location specific information at different levels of granularity. It could be either the very low-level spatial information or the very high-level spatial information and links to concepts in other domains (Barnaghi et al., 2010). Sensors observation and descriptions provided by OGC's SWE Sensor Observation Service (SOS) (Botts et al., 2007) standard is expected to contain location information explained using GML elements. Several dataset such as GeoNames and LinkedGeoData[33] have published an excessive amount of spatial data and named location on the Web as Linked Open Data. Joshua et al. (Pschorr et al., 2010) have presented a semantic sensor network middleware platform that uses and extends the location attribute in OGC's SWE standard for linking it with high-level location concepts and related resources in GeoNames and LinkedGeoData data sources.
- **Temporal Information:** Sensors observations and measurements can be tagged with temporal information describing their time zones and measurements timestamps (Barnaghi et al., 2010).
- **Thematic Information:** Thematic information refers to the real-world high-level events which can be identified using sensors data spatial and temporal information. Thematic data such as types of sensors, tags, features of interest, types of observation measurement, and operational and deployment attributes can be used as a link for describing sensors data in the domain knowledge (Barnaghi et al., 2010).

Semantic annotations, either deduced from the sensors data or physically added by the clients, typically represents information context, having the potential to be fruitfully used in the development of context-aware applications. Context of a resource can be defined by a set of attributes describing situation of the resource depending upon the physical situation including geographic location position (e.g. absolute, relative, and co-location etc), time, infrastructure (e.g. immediate resources for computation, job performance, and communication etc), physical environmental conditions (e.g. noise, light, and pressure etc) (Wei and Barnaghi, 2009). This definition of context more apparently fulfills the Sheth's (Sheth et al., 2008) vision of spatial, temporal, and thematic data. Thus, semantic annotation in a broad level means integration

---

[33] http://linkedgeodata.org/About

of contextual information with sensors observations and measurements.

Semantically annotating sensors data or more generally resources from physical or digital world belonging to the heterogeneous environments will not only help in the exchanging of information and knowledge among different communities in a collaborative environment but will also improves human-machine and machine-machine interactions to ease context-aware applications development and provide considerable advancements in achieving the ultimate goals of ubiquitous computing and Internet of the Future (Wei and Barnaghi, 2009).

5.1.1. Sensor Data Annotation Using Ontologies

Semantic annotation of sensors data provides a mean of relating to highly expressive representations using model references to describe its semantic description. Associating sensors data with semantic description makes data discoverable, accessible, and queryable.

Several languages have been proposed by the research communities for semantically annotating sensors data such as RDFa[34], SAWSDL[35] (Semantic Annotation for SWDL and XML Schema), and Xlink[36] (Sheth et al., 2008). Xlink (XML Linking Language) is a XML markup language for inserting hyperlink elements in XML document to create link between resources belonging to the different XML documents. Xlink created hyperlinks are analogous to HTML but uses XML syntax instead. To provide semantic annotation to sensors data, Xlink attributes can be embedded into SensorML and O&M documents. SWE framework already uses Xlink to semantically annotate SWE documents, therefore, no structural or syntactical changes are required (Compton et al.). RDFa (Resource Description Framework in-attribute) is a collection of attributes which can be used to represent semantic metadata in XML language from which RDF triples can be extracted using simple mapping (Sheth et al., 2008). Like Xlink, RDFa can also be used with SensorML and O&M for semantically annotating sensors data, but requires additional syntax (Compton et al.). SAWSDL (Semantic Annotation for Web Services Description Language) extends Web Services Description Language (WSDL) and XML Schema definition language with a set of attributes for adding semantic description of WSDL components. These attributes annotates WSDL interfaces and operations along with categorization information which can be used while publishing a Web Service in a registry. In addition, SAWSDL can annotate data mapping of XML Schema types to and from ontology. This annotation is independent of the underlying ontology modeling language and does not insist for a particular ontology language. Instead it enables refereeing concepts defined outside of the WSDL document using annotation.

In computer science as well as information science, ontology is a formal representation of the knowledge of a domain, describing the set of concepts as well as the relationships among them within the domain (Breitman et al., 2006). Ontotolgies can be used for annotating sensors data along three dimensions: spatial, temporal, and thematic (Sheth et al., 2008). Several research communities including US National Institute of Standards and Technology (NIST), W3C, and OGC etc have committed their activities towards using ontologies to develop representational models of sensors data. Some of these people have presented novel ontologies, whereas, others have either encompassed or extended the existing onotlogies in the environment in a way. Sensor Standard Harmonization is a project started by NIST for developing a universal sensor ontology using the on hand ontologies in different domains including IEEE 1451[37], OGC SWE languages, ANSI N42.42 [38], and the Chemical, Biological, Radiological, and Nuclear (CBRN) Data Model. W3C Geospatial Incubator Group and OGC Geographic Markup Language are working on building expressive geospatial ontology. W3C recommended a OWL-Time ontology for describing temporal concepts such as instant and interval etc, which can be used in queries emphasizing on time intervals such as contains, within, and overlaps etc. OntoSensor (Russomanno et al., 2005) has extended IEEE Suggested Upper Merged Ontology (SUMO) [39] and incorporated parts of SensorML standard for building ontology-based descriptive specification model for sensors. Eid et al. (Eid et al., 2007) has presented the idea of universal ontology consisting of three individual sub-ontologies: extension plug-in ontology, sensor hierarchy ontology, and sensor data ontology. Payam et al. (Barnaghi et al., 2009) have used the SWE's Observations and Measurements (O&M) as well as SensorML specification to develop Sensor Data Ontology. However, collectively these models do not provide a detail specification of sensors observations and measurements, and do not interpret the complex sensors data as well as the relationships among them.

Domain specific ontologies related to sensor fields such as weather and oceanography etc have been developed providing semantic description of semantic entities (Sheth et al., 2008). Some of the common ontologies belonging to other domains which can be used effectively in annotating sensors data are: (1)

---

[34] http://www.w3.org/TR/rdfa-syntax/

[35] http://www.w3.org/TR/sawsdl/

[36] http://www.w3.org/TR/xlink/

[37] http://ieee1451.nist.gov/intro.htm

[38] http://www.nist.gov/pml/div682/grp04/n42.cfm

[39] http://www.ontologyportal.org/

spatial attributes: SIMILE location ontology [40], DAML location ontology [41]; (2) time attributes: OWL time ontology[42]; (3) common and thematic attributes: CyC ontology [43] and DBpedia ontology [44]. To obtain more information from semantically enriched sensors data, several rule-based reasoning languages such as Semantic Web Rule Language etc have been promoted which allows to define rules on OWL ontologies to infer new semantic asseverations from identified instances (Sheth et al., 2008). SWRL[45] (Semantic Web Language) is a W3C proposed rule-based language for using with OWL ontologies.

Hershal Patni et al. (Patni et al., 2010) have used the concepts defined in observations and measurements (O&M) for defining sensors observations. Sensor ontology is adopted as a schema for modeling and linking sensors data to LOD[46]. In a later work, Hershal Patni et al. (Patni et al., 2010) have used Sensor Provenance (SP) ontology as the core component of the Sensor Provenance Management System (SPMS). SP ontology extends Sensor ontology by incorporating concepts from Provenir upper level ontology. Provenir ontology (Sahoo et al., 2009) is a common modular framework which can be used to create the core components of the provenance management framework. The proposed framework offers a scalable and flexible method to provenance modeling by having the potential to address specific requirements of diverse domains. Using Provenir ontology as a conceptual model for building domain-specific provenance ontologies ensures that: (1) common modeling approach; (2) conceptual clarity of provenance terms; and (c) use of design patterns for consistent provenance modeling. Provenir ontology has re-used and adopted terms, concepts, and properties defined in Relation ontology [47] developed by Open Biomedical Ontologies (OBO) Foundry[48] and has been expressed in OWL-DL[49] with ALCH expressivity. To represent domain-specific provenance information for the sensor domains, Sensor ontology has been extended by incorporating concepts from the Provenir ontology using the rdfs:SubClassOf and rdfs:SubPropertyOf relationships to fashion appropriate classes as well as properties. Classes defined in Sensor ontology related to observation, time, and locations are inherently subclasses of Provenir ontology. Sensor ontology has been described in OWL-DL containing 89 classes, 53 properties, and has ALEHIF+(D) expressivity. As a result SP ontology ensures consistent use of provenance terminology, coherent modeling of concepts, and compatibility with other existing domain-specific provenance ontologies.

Due to extensive adaptation of SWE suite by many people and organizations, Joshua Pschorr et al. (Pschorr et al., 2010) has proposed the extension of OGC’s SWE standards with an ontology for annotating and modeling the sensor descriptions and observations with constructive metadata. Within the SWE O&M standard some properties of an observation might be of complex nature defined in an external document, and thus can be used to describe the relationships of an observation. A new ontology called “O&M-OWL” is developed for encoding observations and measurements in OWL for supporting advanced analysis and reasoning. In addition to previous relations defined, O&M-OWL defines several new relations in a way which can be queried and reasoned effectively for deriving actionable environmental knowledge from sensors observations. To annotate sensors data semantically, semantic terminologies defined within the ontology model is embedded into XML document. Semantic annotation will increase semantic interoperability of sensors data encoded in XML, which inherently only provides synthetic interoperability. Embedding of ontological terms in XML document can be either through reference model or URIs of concepts as defined by the ontology. Using any of the reference concepts to external document defined in the OGC’s SWE standards, a reference model can be embedded into XML document which will provide more meaningful descriptions and ultimately increase semantic interoperability.

5.1.2. Sensor Data Annotation Using Linked Open Data

Linked Open Data itself can be used to semantically annotate sensors data to promote interoperability, establish data standards, and avoid creating the same data repetitively (Wei and Barnaghi, 2009). Data of varied nature already published by several communities on the Semantic Web and used by different projects productively, can also be reused for annotating sensors data. Using Linked Data as a registry for sensors information allows extending existing datasets with new relationships which is a great advantage by its own. Temporal, spatial and thematic concepts already published in Linked Open Data can be used by sensor datasets for semantic annotation with the exception that to extend sensors and observations developed by a publisher requires new relationships to be generated which should refer to the existing facts (Pschorr et al., 2010).

---

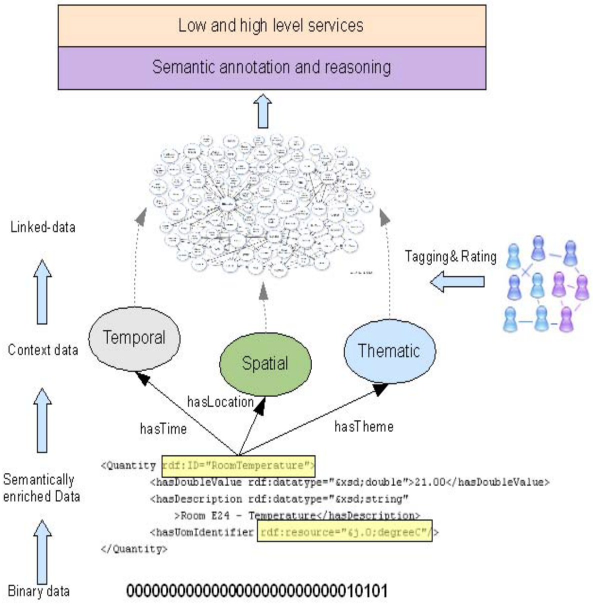
[40] http://simile.mit.edu/2005/05/ontologies/location
[41] http://www.daml.org/experiment/ontology/location-ont
[42] http://www.w3.org/TR/owl-time/
[43] http://cyc.com/cyc/opencyc
[44] http://dbpedia.org/
[45] http://www.w3.org/SWRL
[46] http://knoesis.wright.edu/research/semsci/application_domain/sem_sensor/ont/sensor-observation.owl
[47] http://www.obofoundry.org/ro/
[48] http://www.obofoundry.org/
[49] http://www.w3.org/TR/owl-features/

Wang et al. (Wei and Barnaghi, 2009) has presented the idea of annotating sensors data by using the concepts from the Semantic Web (Linked Open Data) by creating RDF links irrespective of initiating novel classes, occurrences, and users generated contents such as Tags etc. This linking mechanism enables reasoning sensors data as well as Linked Data to present superior sensors data query and retrieval methods. Semantic annotation not only connects sensors data with the Semantic Web but also enriches original or inferred sensors data with abundant of data, increasing the ways of using sensors data productively. In the architecture (shown in Figure 2) local semantically enriched data is annotated spatially, temporally, and thematically with resources from different data sources accessible through LOD. Among others, a reason of Social Web (Gruber, 2008) success is the freedom provided by the Social Websites to the users to annotate easily any object they come through (called Social Tagging). Likewise, in sensor networks the participants are empowered to annotate the sensors data in the same way to exploit “collaborative intelligence”.

5.1.3. Sensor Data Annotation Using Both Ontologies and Linked Data

To use more productively the power and effectiveness of both ontologies and Linked Open Data, a research community has implemented both of them for integrating semantic annotation with sensors data. This method will provide ways of associating a large number of attributes with a sensor data to give it more accurate, specific, and precise description.

Payam Barnaghi et al. (Barnaghi et al., 2010) has proposed a two layer sensors data annotation system: (1) creation of RDF description using a sensor’s main attributes, (2) linking RDF descriptions to existing resource on publically available Linked Open Data. Ontolgoies are used in both steps to include RDF descriptions for representing sensors properties and features. At first basic sensor ontologies, local ontologies, and vocabularies are used for choosing basic terminologies applicable to the sensors from publically available Linked Open Data, providing more specific RDF descriptions, and allowing users to add relevant descriptions to sensor at the time of registering a sensor. Secondly, common ontologies are used to link main properties of sensors data with other RDF files from LOD to provide more specific sensors properties. Publishing such a rich sensor RDF description source will enable creation of resources with few of its properties already defined by other Web resources. It will encourage browsing and accessing a large amount of information about a sensor defined by its different attributes.



```
<Quantity rdf:ID="RoomTemperature">
        <hasDoubleValue rdf:datatype="&xsd;double">21.00</hasDoubleValue>
        <hasDescription rdf:datatype="&xsd;string"
            >Room E24 - Temperature</hasDescription>
        <hasUomIdentifier rdf:resource="&j.0;degreeC"/>
</Quantity>
```

Figure 2. Integrating sensors data with the lined data and Semantic Web (Wei and Barnaghi, 2009)

5.2 Publishing Linked Sensor Data

Collecting a massive amount of data generated by environmental sensors is a common practice of a number of corporate, government, and academic organizations. Sensors made observations about real-world events which are translated into different formats having varying levels of expressivity, according to needs of the applications to exploit the data fruitfully (Patni et al., 2010). Tim Berners Lee[50], coined the idea of LOD with the vital aim of linking and sharing structured information among the people and organizations in the same way as they share documents currently. To fulfill this goal and use sensors data productively, it needs to be openly publically accessible instead of being locked inside organizations. Publishing sensors data refers to the both practices and theories of storing sensors data and related metadata on the Linked Open Data (LOD) for making it openly and publicly accessible (Patni et al., 2010). A normal method to accomplish publishing is a two step process consisting of changing unrefined sensor observations into RDF file and relating RDF file with other data sources on Linked Open Data (LOD). The advent of making a huge quantity of data openly accessible will increase the potential of its utilization and analysis.

Hershal Patni et al. (Patni et al., 2010) have made an effort called “Linked Sensor Data” of integrating sensors domain to LOD. The workflow starts with measuring the environmental phenomena by tons of sensors positioned all round the United States.

---

[50] http://www.w3.org/People/Berners-Lee/

MesoWest[51] service not only aggregates these sensors observations representing the intensities of phenomena but also provides access to them encoded as comma separated numerical values. LinkedSensorData and LinkedObservationData are the two datasets generated where each dataset containing billions of triples. The complete data generation workflow is consisted of four main parts, shown in Figure 3.

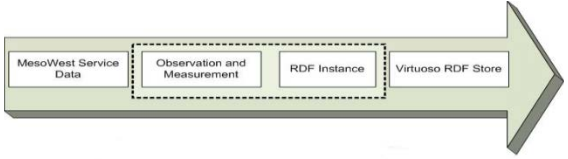


Figure 3. Data generation workflow Linked Sensor Data (Patni et al., 2010).

1. **Phase 1:** The first phase exploits the service provided by MesoWest to query using station ID, data, and time as the parameters. Upon querying MesoWest, a HTML page is returned containing the sensor observations embedded in the weather stations as a comma separated list. Sensor observations data is extracted by parsing the resultant HTML page.
2. **Phase 2:** The extracted sensor observations are feed into SAX (Simple API for XML) parser[52] to convert them into observations and measurements (Botts et al., 2007) (O&M, one of the OGC's Semantic Web Enablement suit of standards and is universally agreed standard for encoding sensors observations by the sensors community).
3. **Phase 3:** Since data encoded in O&M has XML syntax, to annotate data semantically, it needs to be converted into RDF format. Due to their commonality (both O&M and RDF have XML syntax), O&M2RDF-Converter API is being developed using XSLT[53]. O&M2RDF-Converter API is used for converting O&M to RDF.
4. **Phase 4:** In the fourth step, Virtuoso RDF [54] knowledgebase is used to store the generated RDF. Virtuoso RDF is a RDF based triple store developed by Open Link Software and obtainable in open source as well as commercial licenses. Virtuoso RDF is a suite of packages providing command line loaders, an API for connection making, a parser and interpreter for SPARQL, and a web server for executing SPARQL queries plus uploading of information using HTTP. It has been tested to scale up to billion of triples.

Hershal Patni et al. (Patni et al., 2010) in another work has presented Provenance Management Framework (PMF) to implement Sensor Provenance Management System (Sensor PMS). Provenance illustrates the history or the origin of an entity and the data describing the sensor is actually the provenance metadata. Provenance information describes the temporal, spatial, and thematic aspects of a sensor's observations. Adding semantic provenance to define provenance information empowers applications to answer the "why", "where", "what", "when", "which", "who", and "how" queries to truthfully interpret and process data entities. Test beds were conducted using data from 20,000 sensors within the United States in the context to identify blizzard situation. One query was executed to return complete set of provenance information related to a data entity and another query containing a set of constraints defined over the provenance information was executed to return set of data entities satisfying some set of constraints. Results obtained by these two queries were according to the expectations.

Sensor PMS architecture (shown in Figure 4) uses data generation workflow (i.e. identified by (Patni et al., 2010)) with a slight change as the main building block and addresses the three aspects of provenance management suggested by (Sahoo et al., 2009).

- **Provenance Capture:** The provenance information consisting of time related information (temporal information received from MesoWest), and location related information (spatial information retrieved from GeoNames through querying using the sensors coordinates) which are needed to be associated with the sensors are captured during the data workflow.
- **Provenance Representation:** The provenance information related to the sensors are modeled using the specially developed Sensor Provenance (SP) ontology. Sensor Provenance (SP) ontology is developed for adding provenance information within the sensors domain. SP ontology uses the concepts form Provenance Management Framework defined in Provenir upper level ontology to support interoperability among the different provenance ontologies running in the different domains.
- **Provenance Storage:** The modeled provenance information are stored in the Virtuoso RDF store. The Virtuoso RDF store currently contains over a billion triples of sensor observational data and offers a SPARQL endpoint to access datasets.

---

[51] http://mesowest.utah.edu/index.html,
[52] http://www.saxproject.org/
[53] http://www.w3.org/TR/xslt
[54] http://virtuoso.openlinksw.com/dataspace/dav /wiki/Main/VOSRDF

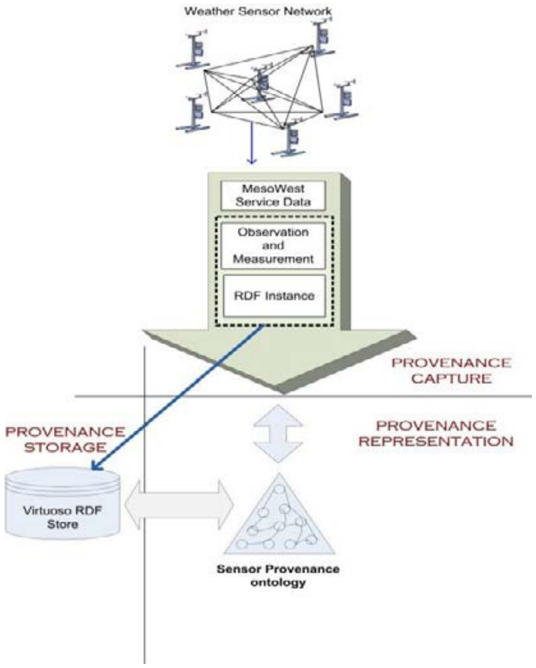


Figure 4. Sensor Provenance Management System architecture (Patni et al., 2010)

Payam Barnaghi et al. (Barnaghi et al., 2010) has presented "Sensor2Web", a system for publishing sensors data using the principles of Linked Data. The system provides environment where users can publish sensors descriptions data in RDF triples, enrich descriptions by associating sensors description data from any defined RDF triple store, interlinking sensors data with elements available in the publically accessible datasets in LOD, and empowering Linked Data consumers to access sensors datasets through SPARQL endpoints.

Sensor2Web architecture's main components are shown in Figure 5. Information related to a sensors such as type, location, and other descriptive material are obtained by querying open linked-data sources (e.g. DBpedia etc) using Jena API[55] and the results are serialized using AJAX technologies directly into the sensors registration web page to help users in filling sensors descriptions. The submitted descriptions are stored in the XML format, which is transformed into RDF form using Extensible Stylesheet Language Transformations (XSLT). With using different stylesheets, the same sensor data can be transformed into different formats or namespaces according to applications needs and requirements. The RDF file contains the main properties of sensors data as well as links to existing RDF files to provide more precise descriptions as per common sensor ontologies. To store the RDF triples SDB[56] (a SPARQL database for Jena) is used. Joseki[57] (a SPARQL server for Jena) is used as SPARQL endpoints. To input more flexibility into the system, several types of interfaces are implemented to provide results to users' queries in different formats such as RDF, XML, and SPARQL protocol format[58].

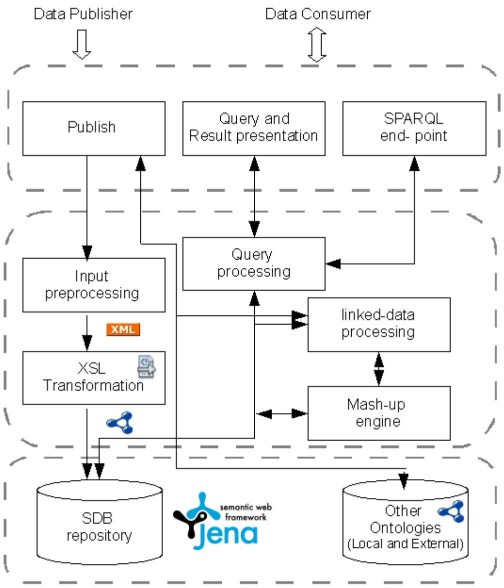


Figure 5. Sensor2Web architecture (Barnaghi et al., 2010)

5.3. Sensors Datasets Description

With the growing popularity and increase in awareness about usefulness of LOD, several research projects, academia, organizations, and individuals have started taking interest in building and publishing data sources on the LOD. In the past few years, LOD has seen a tremendous increase in the number of datasets linked providing a vast knowledge about a number of domains including life sciences, economics, nature, entertainment, and geography etc. Although linking sensors data to LOD is a new paradigm in fact, but still there are some of the online LOD datasets available either providing sensors data and observations or providing real-world information which can be linked and accessed using the data captured by the sensors. In this segment, we describe a summary of such kind of datasets.

5.3.1. Sensors Oriented Datasets

A sensor dataset provides knowledge about sensor measurements and observations in RDF format. At present, LOD cloud has LinkedSensorData and LinkedObservationData as only two datasets, providing sensors data publically accessible.

1. **Linked Sensor Data:** LinkedSensorData is a RDF dataset build by Harshal Patni in the SSW and SST

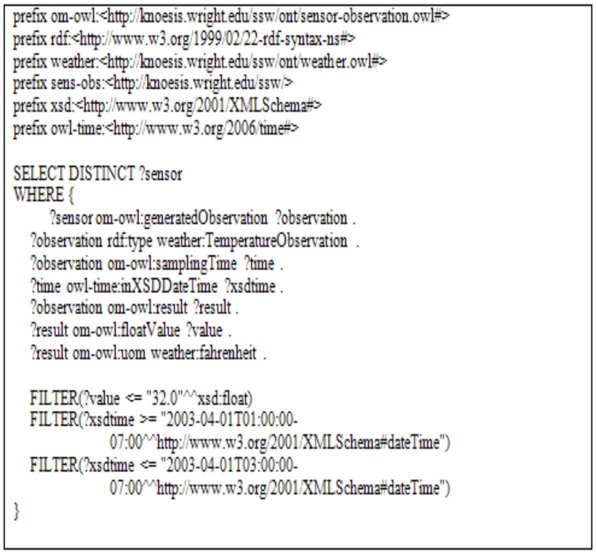
---

[55] http://jena.sourceforge.net/
[56] http://openjena.org/SDB/
[57] http://joseki.sourceforge.net/
[58] http://www.w3.org/TR/rdf-sparql-protocol/

project at Kno.e.sis research center at the Wright State University, USA[59, 60]. This dataset provides expressive descriptions of more than 20 thousands weather stations all around in the United States (Patni et al., 2010)(Patni et al., 2010). The contained data has been originated from the MesoWest project which is collecting weather data since 2002. In MesoWest project, on average five sensors have been deployed per weather station collecting data about a number of phenomena including visibility, temperature, pressure, precipitation, humidity, and wind speed etc. Apart from location attributes such as longitude, latitude, and elevation, the dataset also contains links to locations in GeoNames which are nearby each weather station as well as provide information about the distance from the GeoNames location to the weather station. Apart from them, the dataset also contain links to the most current observations for each weather station provided by MesoWest. This sensor dataset now contains more than 1.7 billion[61] (1730284735) triples and is part of the LOD. The RDF dataset in gzip format is available for download at[62].

2. **Linked Observation Data:** LinkedObservationData is a RDF dataset build by Harshal Patni in the SSW and SST project at Kno.e.sis research center at the Wright State University, USA[52,53]. This dataset contains expressive descriptions of observations about hurricanes and blizzards took place in the United States (Patni et al., 2010)(Patni et al., 2010). The dataset contains observations about several of the major storms which became active after 2002 within the entire United States including Katrina, Hurricane, Bill, Ike, Wilma, Bertha, Gustav, Charley, and a major blizzard in Nevada in 2002. Like LinkedSensorData, LinkedObservationData originates data from the MesoWest project which is collecting weather data since 2002. The observations contained provide information about measurements of phenomena including visibility, temperature, pressure, precipitation, humidity, and wind speed etc. These observations are established using the data generated by weather stations and contained in the LinkedSensorData dataset. At the movement, this dataset contains more than 159 million (159460500) observations and the RDF dataset as well as the statistics for each of the above mentioned storms is available in gzip format for download at[63].

These datasets are provided as SPARQL endpoints, which provide supports for executing SPARQL queries against them for retrieving required data. An example SPARQL query for "*Find all sensors that detected freezing temperatures on April 1, 2003, between 1:00am to 3:00am*" is shown in Figure 6.



```
prefix om-owl:<http://knoesis.wright.edu/ssw/ont/sensor-observation.owl#>
prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix weather:<http://knoesis.wright.edu/ssw/ont/weather.owl#>
prefix sens-obs:<http://knoesis.wright.edu/ssw/>
prefix xsd:<http://www.w3.org/2001/XMLSchema#>
prefix owl-time:<http://www.w3.org/2006/time#>

SELECT DISTINCT ?sensor
WHERE {
      ?sensor om-owl:generatedObservation  ?observation .
   ?observation rdf:type weather:TemperatureObservation  .
   ?observation om-owl:samplingTime  ?time .
   ?time  owl-time:inXSDDateTime  ?xsdtime .
   ?observation om-owl:result  ?result .
   ?result om-owl:floatValue ?value .
   ?result om-owl:uom weather:fahrenheit  .

   FILTER(?value <= "32.0"^^xsd:float)
   FILTER(?xsdtime >= "2003-04-01T01:00:00-
                 07:00^^http://www.w3.org/2001/XMLSchema#dateTime")
   FILTER(?xsdtime <= "2003-04-01T03:00:00-
                 07:00^^http://www.w3.org/2001/XMLSchema#dateTime")
}
```

Figure 6. An example SPARQL query against sensors oriented datasets[64]

5.3.2. Sensors Related Datasets

A wide array of datasets already exists in the LOD cloud which can be linked and accessed using the data provide by the sensors. The data provided by such datasets, if combined with data from sensor datasets can be used in a number of meaningful applications. Such kind of datasets includes repositories of geospatial information (e.g. GeoNames, and Linked Geo Data etc), which are of particular importance in the sensors domain. Some of these datasets are discussed in this section, whereas there are several others which can be potentially exploited by using sensors data.

1. **GeoNames:** Bernard Vatant developed GeoNames[65], which is a geographical database, freely available for download and access using a number of Web services and the database is exported under a Creative Commons Attribution license[66]. Vatant and Mondeca produced the GeoNames ontology[67] for adding geospatial semantic information to the WWW by enriching Linked Open Data Cloud with publishing RDF geographical data (Patni et al., 2010), where each GeoNames toponym has a unique URL with a corresponding RDF Web service. According to estimation, the dataset is currently serving more than 30 million Web services request per day. The dataset currently contains more than 93

---

[59] http://wiki.knoesis.org/index.php/SSW_Datasets
[60] http://wiki.knoesis.org/index.php/Main_Page
[61] http://thedatahub.org/dataset/knoesis-linked-sensor-data
[62] http://wiki.knoesis.org/index.php/SSW_Datasets
[63] http://wiki.knoesis.org/index.php/SSW_Datasets

[64] http://wiki.knoesis.org/index.php/SSW_Datasets
[65] http://thedatahub.org/dataset/geonames-semantic-web
[66] http://www.geonames.org/about.html
[67] http://www.geonames.org/ontology/documentation.html

million triples (93896732)[68], out of which it comprises more than 10 million geographical names, and more than 8 million unique features including 2.8 million populated places and 5.5 million alternate names. All of the features are classified into nine feature classes and are moreover subcategorized into 645 feature codes[69]. The dataset integrates geographical data (e.g. places names etc) in several languages, population, altitude, and more from different repositories. Latitude and Longitude coordinates are maintained in the WGS84 (World Geodetic System 1984) format. A user friendly interface is also provided enabling users to manually add new names or edit, and correct existing names. Third parties have developed SPARQL endpoints for GeoNames dataset.

2. **Linked Geo Data:** LinkedGeoData is an effort of LinkedGeoData (LGD) community project for adding spatial dimensions to the LOD[70]. LinkedGeoData derives data from the comprehensive OpenStreetMap project[71] spatial data collection and uses Linked Data principles to make it available as a large spatial RDF knowledge base for easy to access under the Creative Commons Attribution-ShareALike 2.0 license[72]. LGD consists of over 1 billion node and 100 million ways[73], and the resulting dataset approximately contains 3 billion (3000000000) triples[74]. Data in the LGD is interlinked with other knowledge bases in the LOD like GeoNames, and DBpedia[75] and integrates class labels from translatewiki[76] and icons from the Brian Quinion Icon Collection[77]. The dataset is publically accessible through downloads, SPARQL endpoints, and Linked Open Data[78].
3. **YAGO:** YAGO is a vast semantic knowledge base developed in YAGO-NAGA project at the Max-Planck Institute for Informatics in Saarbrücken/Germany[79]. YAGO knowledge base is constructed by deriving data from WordNet[80], Wikipedia[81], and GeoNames. Currently, YAGO has knowledge about over 10 million entities (e.g. organizations, individuals, and metropolises etc) and possesses over 120 million facts regarding these entities[82]. Unlike other automatically constructed knowledge bases, accuracy of YAGO is manually evaluated having affirmed accuracy of 95%. In YAGO ontology each fact and entity is associated with temporal and spatial dimensions and each relation is annotated with its confidence value. The dataset contains all of the possible names for the entities and is best suited for disambiguation purposes.
4. **US Census:** US Census dataset is developed by Joshua Tauberer by converting 2000 U.S. Census statistics presented by Census Bureau in RDF structure and has made it accessible using SPARQL[83]. The Census data provides population statistics at several geographic levels starting from the U.S at the top, down through states, counties (cities), sub-counties (incorporated towns), so-called "Census Data Places" (villages or towns), ZIP Code Tabulation Areas, and even greater deeper levels of granularity. The statistics provides count information of the total population such as counts by age, sex, and race, information on commuting time to work, mean income, latitude and longitude of the region, etc. Currently, the dataset comprises more that 1 billion RDF triples (1002848918) covering around 3,200 counties, 36,000 towns, 16,000 villages, and 33,000 ZCTAs. The dataset is constructed by first converting Census data into Notation 3 RDF format using a Perl script, and then loaded into MYSQL database, and lastly made it accessible through SPARQL.

6. Sensor Linked Data Applications

Collecting data using sensors, enriching with using Semantic Web technologies, and linking with the Linked Open Data cloud have significant applications in a number of domains (discussed in Section 1) essential for brining facilitation into humans' lives. Programmers with the freedom to access a vast collection of sensors data openly available on the Linked Open Data can find out novel ways of developing useful and productive applications. Using linked sensors data is relatively a novel idea, having great potential but not attracted the attention of application developers due to some of its complexities. Anyhow, researchers have developed some prototypes to leverage the effectiveness of linked sensors data and encourage other developers to join the area.

6.1. Sensor Discovery on Linked Open Data

At present, technology has succeeded in achieving the landmark of deploying millions of sensors around the globe for collecting data about a number of environmental phenomena and publishing

---

[68] http://thedatahub.org/dataset/geonames-semantic-web
[69] http://www.geonames.org/about.html
[70] http://linkedgeodata.org/About
[71] http://www.openstreetmap.org/
[72] http://linkedgeodata.org/Datasets?v=190j
[73] http://linkedgeodata.org/About
[74] http://thedatahub.org/dataset/linkedgeodata
[75] http://wiki.dbpedia.org/Datasets
[76] http://translatewiki.net/wiki/Translating:OpenStreetMap
[77] http://www.sjjb.co.uk/mapicons/
[78] http://blog.aksw.org/2011/linkedgeodata-release-2/
[79] http://thedatahub.org/dataset/yago
[80] http://en.wikipedia.org/wiki/WordNet
[81] http://www.wikipedia.org/
[82] http://www.mpi-inf.mpg.de/yago-naga/yago/
[83] http://www.rdfabout.com/demo/census/

them on the Web to be accessed by a wide verity of useful applications and users (Pschorr et al., 2010). However, finding relevant sensors which might be of potential interest for an application or user on the Web is a non-trivial challenge. For example, sensor located in proximity to a situation, event, or an object is of most importance as compared to the ones father away. Several datasets such as GeoNames and LinkedGeoData etc have published expressive descriptions of spatial data as well as named locations on the Web as Linked Data. Linking sensors descriptions to in close proximity locations expressed by these open datasets will allow issuing sensor discovery queries using named locations.

Sensor discovery is one of the top-priority tasks of W3C Semantic Sensor Network Incubator Group[84] and has tied with the task of developing sensor ontology. As a common consensus by research community, application of Semantic Web technologies can broadly solve the sensor discovery problem. Integrating Sensor Web technologies with Semantic Web technologies will increase the chances of access to the meaningful semantic descriptions of both sensors data and locations provided on Linked Open Data (Pschorr et al., 2010). More precisely, the provenance of sensor observation describing the capabilities of sensors, attributes/properties of sensors, geospatial information of the sensors recording observations (spatial parameters), time stamp of the observations (temporal parameters), phenomena measurement (domain parameters) are essential for answering sensor discovery queries (Patni et al., 2010).

Joshua Pschorr et al. (Pschorr et al., 2010) has presented an idea of semantic sensor network middleware leverages the power of both Semantic Web and existing datasets found on the LOD (e.g. GeoNames and LinkedGeoData etc) for effective discovery of sensors on the Web using named-locations. The authors claim that to use rich, and location-based semantics for sensors discovery, sensors descriptions and observation needs to be annotated with useful metadata. The SWE Sensor Model Language (SensorML) encodes metadata describing the coordinate-based geometric properties of sensors. However, this metadata is enough for determining the geospatial point (i.e. where a sensor is operating etc) but cannot determine the sensors falling within a user’s target location. The proposed idea extends existing SWE framework with integrating Semantic Web technologies by constructing a Semantic Sensor Observations Service (SemSOS). 52North SOS [85] implementation is extended by introducing new SemSOS layer. SemSOS is a set of methods having potential of accessing ontological knowledgebase to support for queries with high-level features such as named-locations etc. Modular architecture of 52North SOS implementation provides support for implementation of SemSOS, where functionalities like request routing, and encoding/decoding etc are kept in place, while the data access implantation at the bottom layer is replaced with SemSOS(as shown as doted box at the bottom layer in Figure 7). The DAOs were replaced with new implementation to provide support for sensor data on Linked Open Data.

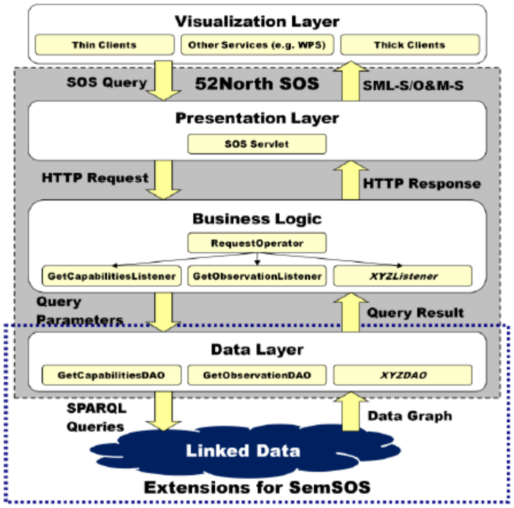


Figure 7. 52North SOS implantation extension with SemSOS (Pschorr et al., 2010).

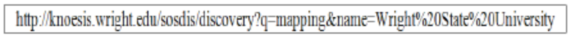
Technically, SemSOS uses the Sesame, and RDF2Go libraries, and SPARQAL queries to access sensor descriptions from LinkedSensorData. SPARQAL queries are generated on the fly by transforming SOS synthetic query parameters (e.g. time, data, and magnitude etc) into triple format confirming to the O&M-OWL ontology. Furthermore, query filters (e.g. location, and comparison operators etc) are converted into SPARQL-type filters and relational operators. The SPARQL queries are executed against triples representing RDF graph and annotated with concepts from O&M-OWL. The resulting graph is then converted into the internal 52North result structure and forwarded to Business Logic layer, where O&M-OWL concepts in the resulting RDF triples are transformed into the original O&M XML encoding. In this way the result from SemSOS client query becomes a valid SOS result.

Using LinkedSensorData as a sensor registry, a prototype SemSOS discovery service has been implemented (Pschorr et al., 2010). To use the service, a user has to input name of location as a feature of interest and the prototype will come up with all sensors and SemSOS services found in the location. Prototype

---

[84] http://www.w3.org/2005/Incubator/ssn/wiki/Main_Page
[85] http://52north.org/communities/sensorweb/sos/index.html

discovery service takes the discovery request in the form of REST (Fielding, 2000) query. An example of the query *"Find me all sensors near the Wright State University"* is given in Figure 8 and an example response to the query in SOS XML format displaying a list of GeoNames features matching the requested query is show in Figure 9.

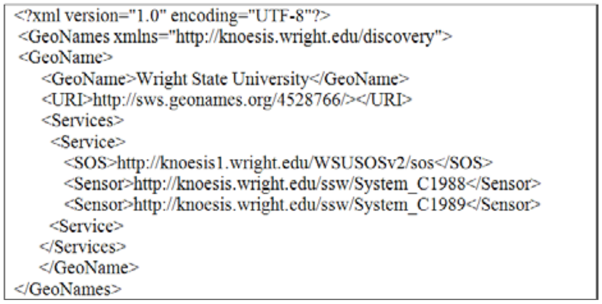


```
http://knoesis.wright.edu/sosdis/discovery?q=mapping&name=Wright%20State%20University
```

Figure 8. Example REST discovery query (Pschorr et al., 2010).

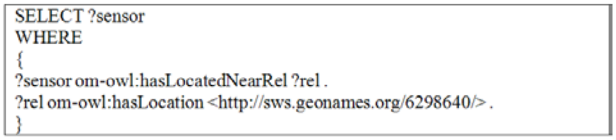
Harshal Patni et al. (Patni et al., 2010) has assumed that linking LinkeSensorData dataset to geographical names provided by GeoNames dataset can be advantageous for answering sensors discovery queries using named-locations such as *"Find all sensors near Dayton-Wright Brothers Airport"*. Example query is shown in Figure 10.

An application with simple map-based GUI has been built for finding nearby sensors using named-location. A user is only required to enter location name in the text box, the application will automatically build and execute SPARQL query over the LinkedSensorData dataset on LOD and renders all of the sensors available nearby the given location on a map. In addition to simply displaying sensors on the map, other information related to the geographic conditions, sensor description and phenomena measured are also displayed. Sensor description contains links pointing to the original data sources such as current sensor data on MesoWest, and location information on GeoNames.



```
<?xml version="1.0" encoding="UTF-8"?>
 <GeoNames xmlns="http://knoesis.wright.edu/discovery">
 <GeoName>
    <GeoName>Wright State University</GeoName>
    <URI>http://sws.geonames.org/4528766/></URI>
    <Services>
     <Service>
       <SOS>http://knoesis1.wright.edu/WSUSOSv2/sos</SOS>
       <Sensor>http://knoesis.wright.edu/ssw/System_C1988</Sensor>
       <Sensor>http://knoesis.wright.edu/ssw/System_C1989</Sensor>
     <Service>
    </Services>
    </GeoName>
</GeoNames>
```

Figure 9. Example SOS XML-based discovery response (Pschorr et al., 2010).



```
SELECT ?sensor
WHERE
{
?sensor om-owl:hasLocatedNearRel ?rel .
?rel om-owl:hasLocation <http://sws.geonames.org/6298640/> .
}
```

Figure 10. Example SPARQL sensor discovery query (Patni et al., 2010).

6.2. Sensor Data Mashups

Danh Le-Phuoc et al. (Phuoc and Hauswirth, 2009) has presented a system called "SensorMasher" which integrates sensors data available on LOD into mashups. SensorMashor composes and enables non-technology oriented users to derive new sensors data sources by fusing existing sensors data from multiple sources which are identified by URIs and are integral parts of virtual RDF graph in a visual way. The flow of data from sensors to data stream and from data stream to data processing operators are controlled by the mashup's configurations which are generated by the composing process and stored in the metadata repository. The triple-based query processor will solve the problem of exploring the data sources of interest among the existing ones using semantic-based sensor data exploration. Faceted browser will not only enable users to navigate from one dataset to another by following the semantic links but will also help in filtering sensor data using relevant facts. SensorMasher also provides SPARQAL endpoint, allowing users to post all types of queries (e.g. identifying a sensor or retrieving all types of sensors data etc) using semantic based descriptions of sensors to this SPARQL endpoint. All types of queries including exploration and navigation are transformed into complex queries to the query processor under the hood.

Architecture of SensorMasher presented by (Phuoc and Hauswirth, 2009) is show in Figure 12. The wrappers at the bottom interfaces with the physical sensors (i.e. using USB, serial ports, and Bluetooth etc) for receiving stream of sensor data in both push-based and pull-based fashions. Sensor data arriving at the stream would not be in suitable format to be handled directly in persistent storage system like relational database for a number of reasons. Therefore, Data Stream Management System (DSMS) is responsible for monitoring, analyzing, combining, and correlating streams of data, instead of directly following the methods of traditional data management systems. DSMS not only provides interfaces for controlling sensor data streams but also provides APIs for supporting regular queries over the data streams. Once the steam is created and data arrives on the stream, the DSMS forwards the data to fusion operators. The fusion operators uses the methodologies form the multi-sensor data fusion process models for performing data processing operations including filtering, alignment, association, correlation, and classification etc on sensors observed data. A fusion operator also extracts the semantic from sensor readings added by the wrappers. Sensor & Mashup manager component is responsible for monitoring and controlling the whole process of data flow among wrappers, DSMS, and fusion operators. The Sensor & Mashup manager also provides interfaces for the Mashup Composer and Web Interfaces for performing several operations including editing and querying the metadata etc in the triple-based model. It also interacts with the User Manager component responsible for managing the mashups' authors' profiles and mashups' sharing policies. The query processor component executes the continuous queries over the DSMS component and the triple-based queries over the metadata data source. The Explorer component is a simple GUI for clients to investigate sensor data as well as semantic links on the map and also include facet-based features for filtering over the triple-based sensor data. The Web interface includes a SPARQL endpoint, Web services interface, and receives HTTP requests from URL-based Sensor Web sources. The Mashup Composer component allows users to visually compose sensor mashups by connecting sensor data sources with fusion operators.

SensorMasher is implanted and available at *sensormasher.deri.org*. Initially SQL was used to support declarative continuous queries. Afterword, SPARQL query processor was built by extending the Jena ARQ[86] using the approach of D2RQ (Bizer and Cyganiak, 2006) on top of the both DSMS and metadata repository. The mapping rules are automatically generated from the configuration of mashups. To control the SensorMasher two ontologies: core and extended ontolgoies are used. Core ontology describes classes and attributes which are similar for each SensorMasher deployment but extended ontology contain customized subclasses of core ontology according to specific requirements. Extended ontology is created using the classes and attributes from SWEET property ontology and the SANY sensor taxonomy. While answering a SPARQL query, Jena-in-memory reasoner is used to generate query mapping rules using these two ontologies at the class level. The Explorer and Mashup Composer are implemented as AJAX-based web applications.

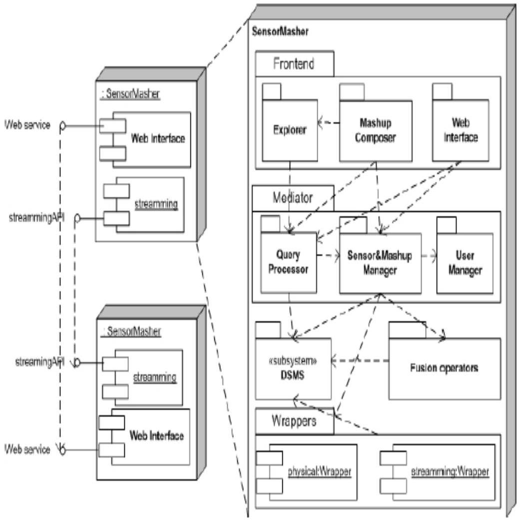


Figure 12. Architectural overview of SensorMasher (Phuoc and Hauswirth, 2009).

Payam Bernaghi et al. (Barnaghi et al., 2010) has also demonstrated a mashup application using Google Maps API to show the Linked Open Data utilization and assimilation of data from multiple repositories. A user has to only provide location attribute of a resource. The application then extracts geographical coordinates as well as other related attributes of the resource from the Linked Open Data and show existing sensors along with their attributes using Google Map application. Currently, the application only retrieves published properties on the map but can be improved to discover other related resource (e.g. nearby locations, and objects etc) through following different links. To retrieve more information about sensors by linking properties of sensors with other resources would require availability of common ontologies describing diverse aspects of sensors such as platforms, types, measurement attributes, and devices etc.

7. Conclusion and Recommendations

Linking sensors data to Linked Open Data cloud is a working solution to make sensors observations and measurements openly available and provides an environment to integrate sensors data with data contained in the existing data sources on the Semantic Web in an easy way. For clear and efficient understanding of linking sensors data to the Linked Open Data, we comprehensively surveyed the available literature in this area. Form the existing literature, it has been observed that integrating phenomenon from physical world to data already existing on the digital world can be used in the design of wide range of useful front-end applications, services, and systems such as health, manufacturing, monitoring, tracking, and planning. Semantic Web technologies can be used to solve the issues of heterogeneous sensors data representation, annotation, sharing, management, and reasoning of sensors data. Apart from them, publishing sensors data on Linked Open Data using Semantic Web technologies from heterogeneous domains enables dissimilar communities to communicate information and knowledge in a collaborative environment.

Lack of infrastructure is not the primary obstacle in the path of linking sensor data, as millions of sensors are either deployed around the world or available in the networked embedded devices. But, there are some technological shortcomings which needs to be addressed to exploit the full spirit of sensors data linked to Linked Open Data cloud. After a comprehensive survey, we came up with a number of recommendations which needs immediate attention of the research communities, organizations, and academia. These recommendations are:

---

[86] Jena ARQ - http://jena.sourceforge.net/

- As most of the research work in the area is contributed by the research projects, organizations, and academia. To bring people and programmers with lack of knowledge about Linked Open Data and Semantic Web into the loop, specialized applications such as visual composers and explorer needs to be developed to help them in deploying, combining, annotating, linking, and searching sensors data sources.
- To speed up and facilitate the conversion of sensors data produced in heterogeneous formats into RDF or OWL, new specialized and easy-to-use tools needs to be developed or problems associated with existing tools needs to be resolved. For example, JXML2OWL suffers from accurately mapping of sensor phenomenon to appropriate classes in the underlying OWL ontology, and XML2OWL suffers with lack of Web Service and API.
- The ubiquitous nature of sensors and their capabilities to measure data from users' environment and publish them on the Web of Data might have put the users' security and privacy at stack. Therefore, specialized measures needs to be researched and developed to ensure users' security and privacy in linking sensor data scenarios.
- The spatio-temporal dynamics of sensors and their corresponding observation data can create serious problems. Therefore, appropriate representations in RDF needs to be discovered to accurately annotate spatio-temporal properties with timestamps and locations.
- Instead of creating domain specific ontologies or schemas for mapping and modeling heterogeneous sensor data in a particular domain, a comprehensive sensor ontology needed to be investigated which should provide deep sensors knowledge model instead of capturing only superficial sensors attributes. The sensor ontology should have the potential to address all types of sensors, their observations, and applicable in all possible domains.
- In the sensors domain, a phenomenon is an effect caused by a number of possible features, events, and real-world objects. Adductive reasoning engine needs to be incorporated to reason from sensors observations of phenomena to identify possible hypothesis.
- Instead of developing additional ontologies for modeling heterogeneous sensor data, it is also possible that the addition of semantics to all of the OGC's SWE standards (i.e. instead of only SOS) will provide an improved platform for finding, using, controlling, and reasoning over sensor data as well as observations on the Web.
- Instead of only finding location of fixed-location sensors on the Web, techniques are needed to be developed to find locations of mobile sensors on the Web as well.
- With the significant importance of sensors data on LOD, simple GUI based applications needs to be developed to help programmers in leveraging the strength of linked sensors data such as Web mashup developers can benefit by using URIs to access sensors observations encoded in accepted formats such as RDF etc.
- Efficient storage mechanism needs to be developed to store enormous amount of data generated either by the sensors continuous sensing, by the queries defined over a period of time, or archiving purposes. Experimental results have show that currently available triple storage cannot handle efficiently high volume of data especially in searching and updating cases.

We believe that this effort will provide help and guidelines to the researchers for understanding and instigating further research on this demanding topic, and will encourage system and application developers to build valuable and appealing linked sensor data applications to make the dream "Web of Data" true.

References

1. Balazinska M, Deshpande A, Franklin MJ, Gibbons PB, Gray J, Hansen M, Liebhold M, Nath S, Szalay A, Tao V. Data Management in the Worldwide Sensor Web. IEEE Pervasive Computing, 2007:6(2): 30-40.
2. Barnaghi P, Meissner S, Presser M, Moessner K. Sense and Sens'ability: Semantic Data Modelling for Sensor Networks. In proccedings of the ICT Mobile Summit 2009, Santander, Spain, 2009.
3. Barnaghi P, Presser M, Moessner K. Publishing Linked Sensor Data. In proceedings of the 3rd International Workshop on Semantic Sensor Networks (SSN), Organised in conjunction with the International Semantic Web Conference (ISWC) 2010, Shanghai, China, 2010.
4. Bizer C, Cyganiak R. D2R Server--publishing relational databases on the semantic web. Poster at the 5th International Semantic Web Conference, 2006.
5. Bizer C, Heath T, Berners-Lee T. Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems (IJSWIS), 2009:5(3): 1-22.
6. Botts M, Percivall G, Reed C, Davidson J. OGC Sensor Web Enablement: Overview And High Level Architecture. OGC White Paper, 2007: OGC 07-165 (45/40/2008):175-190.
7. Breitman K, Casanova MA,Truszkowski W. Semantic Web: Concepts, Technologies and Applications (NASA Monographs in Systems and

Software Engineering). Springer-Verlag, New York, 2006,

8. Compton M, Henson C, Neuhaus H, Lefort L, Sheth A. A Survey of the Semantic Specification of Sensors. In proceedings of the 2nd International Workshop on Semantic Sensor Networks (SSN09), Washington DC, USA, 2009.
9. Corcho O, García-Castro R. Five challenges for the Semantic Sensor Web. Semant. web, 2010:_1(1,2): 121-125.
10. Eid M, Liscano R, El Saddik A. A Universal Ontology for Sensor Networks Data. In proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems and Applications 2007 (CIMSA 2007), 59-62, 2007.
11. Fielding RT. Architectural styles and the design of network-based software architectures. University of California, Irvine. 162, 2000.
12. Gruber T. Collective knowledge systems: Where the Social Web meets the Semantic Web. Web Semant., 2008:6(1): 4-13.
13. Keßler C, Janowicz K. Linking Sensor Data – Why, to What, and How?. In proceedings of the 3rd International workshop on Semantic Sensor Networks 2010 (SSN10) in conjunction with the 9th International Semantic Web Conference (ISWC 2010), Shanghai, China, 2010.
14. Le-Phuoc D, Parreira JX, Hausenblas M, Han Y. Live linked open sensor database. In proceedings of the 6th International Conference on Semantic Systems, Graz, Austria, 1-4, 2010.
15. Le Phuoc D. SensorMasher: publishing and building mashup of sensor data. Retrieved: 23/01/2013, 2013, From: http://triplify.org/files/challenge_2009/sensormasher.pdf.
16. Leggieri M, Passant A, Hauswirth M. inContext Sensing: LOD augmented sensor data. In proceedings of the 10th International Semantic Web Conference ISWC2011, 2011.
17. Patni H, Henson C, Sheth A. Linked Sensor Data. In proceedings of 2010 International Symposium on Collaborative Technologies and Systems (CTS 2010), Chicago, USA, 2010.
18. Patni H, Sahoo SS, Henson C, Sheth A. Provenance Aware Linked Sensor Data. In proceedings of the 2nd Workshop on Trust and Privacy on the Social and Semantic Web, Co-located with ESWC, Heraklion, Greece, 2010.
19. Phuoc DL, Hauswirth M. Linked Open Data in Sensor Data Mashups. In proceedings of the 2nd International Workshop on Semantic Sensor Networks (SSN09), in conjunction with ISWC 2009, 2009.
20. Pschorr J, Henson C, Patni H, Sheth A. Sensor Discovery on Linked Data. Kno.e.sis Center Technical Report, 2010.
21. Raskino M, Fenn J, Linden A. Extracting Value From the Massively Connected World of 2015. Gartner Research Technical Report, 2005.
22. Russomanno DJ, Kothari C, Thomas O. Sensor Ontologies: From Shallow to Deep Models. In proceedings of the IEEE Thirty-Seventh Southeastern Symposium on System Theory 2005 (SSST '05), 107-112, 2005.
23. Sahoo S, Barga RS, Goldstein J, Sheth A, Thirunarayan K. Where did you come from...Where did you go? An Algebra and RDF Query Engine for Provenance. Kno.e.sis Center, Wright State University, 2009.
24. Sahoo S, Weatherly DB, Mutharaju R, Anantharam P, Sheth A,Tarleton R. Ontology-Driven Provenance Management in eScience: An Application in Parasite Research. On the Move to Meaningful Internet Systems: OTM 2009, Springer Berlin Heidelberg, 5871, 992-1009, 2009.
25. Sheth A, Henson C, Sahoo SS. Semantic Sensor Web. Internet Computing, IEEE, 2008:12(4): 78-83.
26. Wei W, Barnaghi P. Semantic annotation and reasoning for sensor data. In proceedings of the 4th European conference on Smart sensing and context, Guildford, UK, 66-76, 2009.

10/12/2013