# Semantic Enrichment of Vital Sign Streams through Ontology-based Context Modeling using Linked Data Approach

Sachiko Lim, Rahim Rahmani and Paul Johannesson
*Department of Computer and Systems Science, Stockholm University, Kista, Sweden*

Keywords: Internet of Things (IoT), Semantic Enrichment, Ontology, Linked Data, Patient Health Monitoring, Patient Management, Vital Sign, Healthcare, Infectious Disease Outbreak.

Abstract: The Internet of Things (IoT) creates an ecosystem that connects people and objects through the internet. IoT-enabled healthcare has revolutionized healthcare delivery by moving toward a more pervasive, patient-centered, and preventive care model. In the ongoing COVID-19 pandemic, it has also shown a great potential for effective remote patient health monitoring and management, which leads to preventing straining the healthcare system. Nevertheless, due to the heterogeneity of data sources and technologies, IoT-enabled healthcare systems often operate in vertical silos, hampering interoperability across different systems. Consequently, such sensory data are rarely shared nor integrated, which can undermine the full potential of IoT-enabled healthcare. Applying semantic technologies to IoT is a promising approach for fulfilling heterogeneity, contextualization, and situation-awareness requirements for real-time healthcare solutions. However, the enrichment of sensor streams has been under-explored in the existing literature. There is also a need for an ontology that enables effective patient health monitoring and management during infectious disease outbreaks. This study, therefore, aims to extend the existing ontology to allow patient health monitoring for the prevention, early detection, and mitigation of patient deterioration. We evaluated the extended ontology using competency questions and illustrated a proof-of-concept of ontology-based semantic representation of vital sign streams.



## 1 INTRODUCTION

Healthcare has marked a significant paradigm shift from a centralized, professional-focused, and reactive model to a more pervasive, patient-centered, and preventive care model (Epstein et al., 2010). The Internet of Things (IoT) creates an ecosystem that connects people and objects through the internet. By revolutionizing healthcare service delivery, IoT-enabled healthcare has a high potential to improve population health and transform a healthcare model toward a more comprehensive personalized care model (Kelly et al., 2020). For example, it enables preventive primary healthcare services to be more accessible and available by enabling remote and real-time monitoring of the patient's health status and daily activities, leading to a more proactive prediction of health issues (Kelly et al., 2020).

The current COVID-19 pandemic has posed devastating effects on global health and economies. Given that the incidence of emerging infectious diseases has been increasing at an unprecedented rate (Jones et al., 2008), it also has manifested the pressing need for more resilient healthcare systems against future emerging infectious diseases. IoT technology has been exerting its power through its potential to mitigate the impacts of COVID-19 on individuals and health systems. Those innovative IoT have been used for screening and early diagnosis of COVID-19, triage of patients, epidemiological surveillance (Golinelli et al., 2020), and contact tracing (Swayamsiddha & Mohanty, 2020). Other utility includes maintaining social distancing, remote and real-time monitoring of confirmed/asymptomatic/suspected cases, ensuring adherence to isolation/quarantine, and after-recovery follow-ups to understand long-time sequelae and possible re-infection (Nasajpour et al., 2020). By providing such capabilities, IoT-enabled healthcare systems can also help prevent overstretching healthcare systems (Swayamsiddha & Mohanty, 2020). Thus, the successful integration of IoT technology in existing healthcare systems seems to be a key to increase preparedness and resilience for future pandemics.



Lim, S., Rahmani, R. and Johannesson, P.
Semantic Enrichment of Vital Sign Streams through Ontology-based Context Modeling using Linked Data Approach.
DOI: 10.5220/0010582202920299
In *Proceedings of the 10th International Conference on Data Science, Technology and Applications (DATA 2021)*, pages 292-299
ISBN: 978-989-758-521-0

A dramatic increase in IoT usage has generated a massive amount of heterogeneous sensory data. Nevertheless, due to the heterogeneity of data sources and technologies, IoT-enabled healthcare systems are often operating in vertical silos, which hampers interoperability across different systems (Kelly et al., 2020). As a result, such sensory data are rarely shared nor integrated. Moreover, a previous systematic literature review identified contextualization and situation-awareness as some of the challenges IoT applications in healthcare have been facing (Lim & Rahmani, 2020). An underlying reason for the issue is that raw sensory data are mere numerical values that are not necessarily easy to associate with meaningful and understandable information unless the context of data is provided (Ganz et al., 2016).

Applying semantic web technologies to represent IoT data is a promising approach for fulfilling heterogeneity, contextualization, and situation-awareness requirements for real-time healthcare solutions. Adding semantic description can transform raw data into an unambiguous machine-interpretable form. The semantically enriched data can further be processed and interpreted by machines to generate meaningful knowledge (Biswanath, 2017).

The need for and execution of IoT-enabled healthcare services largely varies depending on the context (Alirezaie et al., 2017). Ontologies are among the most appropriate approaches to perform context modeling (Strang & Linnhoff-Popien, 2004). Ontology refers to "a formal naming and definition of the types, properties, and relationships of the entities that really or fundamentally exist in a particular domain of discourse" (Biswanath, 2017). Ontologies are machine-interpretable due to the formal and explicit specification of conceptualizations, which has capabilities of knowledge sharing, logic inferencing, knowledge reuse, and knowledge integration (Biswanath, 2017; Perera et al., 2014).

Linked Data is another critical pillar of semantic technologies, which refers to "a set of best practices for publishing and interlinking structured data on the Web." Linked Data connect items across different data sources in a single global data space (Heath & Bizer, 2011). Using an ontology complements Linked Data by supporting data integration, schema alignment, reasoning, and inferencing over data (Biswanath, 2017).

Relevant previous studies have mainly focused on annotating cross-sectional/categorical data rather than continuous data. Jabbar et al. proposed an IoT-based Semantic Interoperability Model (IoT-SIM) to improve semantic interoperability among heterogeneous IoT devices in the healthcare domain (Jabbar et al., 2017). In their study, physicians made a diagnosis using IoT devices and prescribed medicine accordingly. They then semantically annotated the diagnosis and prescription results (i.e., cross-sectional data) in RDF. Carbonaro et al. proposed an ontology-based cognitive computing eHealth system to achieve semantic interoperability among heterogeneous IoT fitness and wellness applications (Carbonaro et al., 2018). However, they did not describe any details about the steps to annotate sensor data.

There is, thus, a lack of studies that have performed the enrichment of the sensor data streams with their spatial, temporal, semantic meaning (Pacha et al., 2020). Furthermore, to our knowledge, there is no existing ontology enabling continuous patient health monitoring for more effective patient management during infectious disease outbreaks.

Therefore, this study aims to extend the IoT-Stream ontology (Elsaleh et al., 2020) in order to enable patient health monitoring for the prevention, early detection, and mitigation of patient deterioration. Using the extended ontology, we performed ontology-based context modeling of vital signs (i.e., mapping vital sign data to ontology concepts) to add contextual information to raw sensor data (i.e., semantic enrichment). We formulated the following research question: "How can ontology-based context modeling be used for the semantic representation of vital sign streams from heterogeneous data sources for enabling patient health monitoring and management?" To address the research question, we performed the following four steps:

1) Extend the existing ontology, named the IoT for patient health monitoring (IoT4PHM) ontology, to enable patient health monitoring and management.
2) Extract vital signs data from two different data sources.
3) Build and apply a semantic model to semantically enrich vital signs using the Resource Description Framework (RDF).
4) Perform a semantic search using a SPARQL query language.

## 2 RELATED WORK

### 2.1 Ontology-based Context Modeling of Sensor Data

Semantic Sensor Network (SSN) is an OWL 2 ontology that is a de-facto standard ontology for describing sensors, their accuracy and capabilities, observations, and sensing methods (Compton et al., 2012).

Nevertheless, context modeling using heavy-weight ontologies can increase computational complexity and processing time, especially when data volume is high (Perera et al., 2014). Such heavy-weight ontologies are thus not suitable for (near) real-time IoT applications. Light-weight ontologies, therefore, have been created to cope with those real-time applications.

For example, the Sensor, Observation, Sample, and Actuator (SOSA) ontology is a light-weight, general-purpose ontology that provides a flexible but coherent framework to represent the entities, relations, and activities involved in sensing, sampling, and actuation (Janowicz et al., 2019). IoT-Lite is another light-weight semantic model for IoT (Bermudez-Edo et al., 2017). The ontology instantiates the SSN ontology to describe key IoT concepts, enabling interoperability and discovery of sensory data among heterogeneous IoT platforms. However, SOSA and IoT-Lite focus on devices rather than IoT data streams (Elsaleh et al., 2020). Thus, IoT-Stream, a light-weight semantic model for annotating IoT streaming data, was created by extending SOSA.

### 2.2 Semantic Enrichment of Sensor Data Streams

Discovering and analyzing sensor data requires spatial, temporal, and thematic information. Nevertheless, sensor observation data are by nature opaque. Thus, metadata is crucial for managing sensor data. A semantic sensor web (SSW) enriches sensor data by providing the meaning for the sensor data. The semantic enrichment facilitates interoperability and enables situational awareness and advanced application from heterogeneous sensors (Sheth et al., 2008).

As written in Section 1, few studies have focused on annotating sensor data streams. However, there are still several pioneering studies that aimed to address the research gap. Pacha et al. proposed a novel framework called SEmantic Annotation over Summarized sensOr Data stReam (SEASOR), enabling the real-time semantic annotations of streaming sensor data (Pacha et al., 2020). Their framework facilitates sensor data stream analytics through summarization, semantic annotation, and query processing (Pacha et al., 2020). Semantic annotation is performed over the summarized sensor data using the base ontology extended from the SSN ontology.

Alirezaire et al. presented a system called E-care@home, which can integrate measurements from heterogeneous sensor sources used for ambient assisted living using ontologies (Alirezaie et al., 2017). By augmenting devices and their measurements into semantic representation, the system provides the semantic interpretation of events and context awareness. To semantically represent sensor data, they developed the SmartHome ontology, which includes modules representing a smart home environment's physical and conceptual aspects. They proposed to use a network of interlinked ontology modules because real-time reasoning requires reducing the complexity of semantic reasoning (Alirezaie et al., 2017).

## 3 METHODS

### 3.1 Data Preparation

#### 3.1.1 Electronic Medical Record (EMR) Vital Sign Dataset

We extracted patient vital signs from the early prediction of sepsis from clinical data published on the PhysioNet website (Reyna et al., 2020). The dataset consists of hourly vital sign summaries, laboratory values, and static descriptions of 60,000 ICU patients from two hospitals. It includes 40 clinical variables: 8 vital signs, 26 laboratory variables, and six demographic variables (Reyna et al., 2020). Of the 40 variables, this study utilized the following seven variables: age, gender, heart rate, oxygen saturation, temperature, systolic blood pressure, and respiration rate.

In addition, we simulated sensor ID by random number generation and the start and end time of the IoT stream, assuming that the selected vital signs are monitored simultaneously. Since the location data are required for context-aware health systems, we conveniently extracted the location data from the epidemiological dataset of the COVID-19 outbreak (Xu et al., 2020). We randomly selected ten patients who developed sepsis within 10 hours after admission to ICU and used their 10-hour vital sign observations for the semantic enrichment.

#### 3.1.2 Radar Vital Sign Dataset

In the EMR vital sign dataset, every vital sign is provided in hourly summaries. To demonstrate the semantic enrichment of raw sensor data, we extracted ECG raw data from a publicly available dataset which consists of 24 h of synchronized data from radar and a reference device (Schellenberger et al., 2020). The dataset contains data including ECG, impedance cardiogram, and non-invasive continuous blood pressure collected from 30 healthy participants. We extracted ECG signals, tfm_ecg1, which were recorded at the sampling frequency of 2000 Hz.

First, we randomly selected 10 of the total 30 participants and assessed their ECG signals' length. Since one participant had the shortest ECG length of 1200000 (corresponding 10 minutes recording), we truncated other participants' ECG to this point so that every participant has the same ECG length.

We then applied a sliding window for the real-time processing of ECG. We set the size of the sliding window to 150000 samples (i.e., 60 seconds) and the step size of the sliding window to 120000 samples (i.e., 60 seconds), which indicates the overlapped interval between sliding windows is 30000 samples (i.e., 15 seconds).

After setting the sliding window, step size, and overlapped interval, we applied the Pan Tompkins algorithm for each sliding window using the MATLAB function, "Complete Pan Tompkins Implementation ECG QRS detector" (Sedghamiz, n.d.). The algorithm is most widely used to detect QRS complex for detecting and monitoring various cardiovascular diseases (Fariha et al., 2020). After detecting R peaks in each sliding window, RR intervals were determined to compute heartbeats.

Please note that optimizing R-peak detection and the size of sliding window and step size is out of this study's scope. Other preprocessing methods for denoising ECG and detecting QRS complexes can undoubtedly be used to achieve clinical relevance.

We prepared the datasets from both data sources with MATLAB ver. R2020b.

## 3.2 Context Modeling of Vital Signs using Linked Data

We chose ontology-based context modeling because the approach was identified to be the most promising asset for context modeling in ubiquitous computing (Strang & Linnhoff-Popien, 2004). Figure 1 shows the overview of the IoT4PHM ontology. We reused the IoT-Stream ontology. The ontology focused on modeling an IoT stream, stream observations belonging to the IoT stream, and analysis used for and events detected from the IoT stream. Those concepts are captured in four classes: IoTStream, StreamObservation, Analytics, and Event, depicted in sky blue rectangles in Figure 1 (Elsaleh et al., 2020). The IoTStream class is the central concept of the ontology, representing an IoT data stream generated by an IoT source. The StreamObservation class is continuous stream observations belonging to the IoT stream, observed by a sensor device captured as a data point over a time instant or a subset of data points over a defined time interval. The Analytic class captures the data analytics that has been applied to analyze the IoT data stream. The Event class abstracts the event that has been detected by using an analytics process to an IoT data stream (Elsaleh et al., 2020).

Moreover, the IoT-Stream ontology is linked with six concepts from external ontologies (Figure 1): qoi:Quality, iot-lite:Service, sosa:Sensor, qu:QuanityKind, qu:Unit, and geo:Point. The qoi:Quality is the top class to describe the quality of IoT data sources. The class has a sub-class called Timeliness which defines a metrics category to represent which rate a data source provides data within a defined time span or age. The iot-lite:Service is an abstract of a service provided by an IoT device. The sosa:Sensor is a device, agent (including humans), or software (simulation) that generates an IoT stream. The qu:QuanityKind represents a quantity without any numerical value or unit, while qu:Unit abstracts the concept of measurement unit. The geo:Point represents the latitude, longitude, and altitude of the location where an IoT stream originates. For a complete description of the ontology, see (Elsaleh et al., 2020).

We added three concepts (Figure 1) to support the data integration and sharing, knowledge representation and reasoning, and computer-assisted data analysis to enable patient health monitoring and management for the prevention, early detection, and mitigation of patient deterioration.

The first concept, Patient, stores the patient-related information such as ID, age, sex, and social determinants of health, which can have a considerable effect on health outcomes. For example, the increasing evidence shows that social determinants of health, including poverty, physical environment, race, or ethnicity, impacts COVID-19 morbidity and mortality profoundly and unevenly (Abrams & Szefler, 2020). Thus, the concept is essential for understanding the patient's needs to provide optimal healthcare services.

The second concept, UnderlyingHealthCondition, is added to understand the patient's underlying health conditions because they can significantly increase the risk of a worse disease prognosis. For example, a modeling study shows that the population with underlying health conditions such as chronic kidney

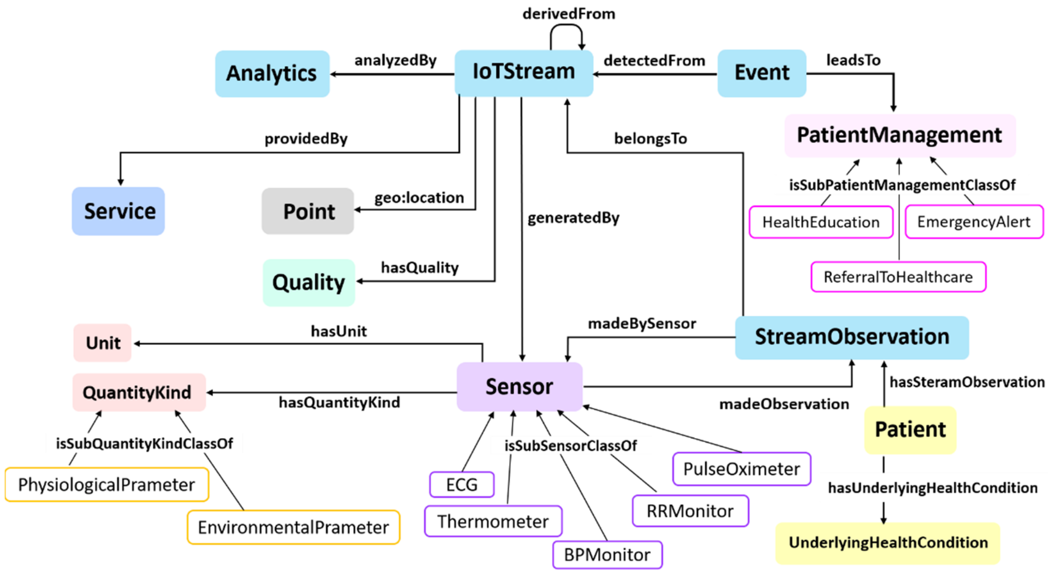


Figure 1: The overview of the IoT4PHM ontology.

disease, diabetes, cardiovascular disease, and chronic respiratory disease are at increased risk of severe COVID-19 and hospitalization (Clark et al., 2020). Having the UnderlyingHealthCondition class enables identifying high-risk groups, which is crucial for performing triage and rolling out effective epidemic management (e.g., identifying individuals who may need to be shielded or vaccinated first).

The third class, PatientManagement, is added because it is essential to enable a computer-assisted analysis to recommend patient management measures (e.g., perform semantic reasoning to compute early warning score and recommend the corresponding patient management measure). The class has three subclasses: HealthEducation, EmergencyAlert, and ReferralToHealthcare. The HealthEducation represents the concept of general public health practices recommended by, for example, a public health agency to reduce the transmission of infectious diseases. The EmergencyAlert class abstracts a concept of clinical alert that can be critically important for an individual (e.g., dispatch an ambulance). The ReferralToHealthcare represents the concept of referring the patient to healthcare.

We also added PhysiologicalParameter and EnvironmentalParameter as subclasses of QuantityKind, which refers to an "aspect common to mutually comparable quantities" and represents the essence of a quantity without any numerical value or unit (e.g., humidity) (Elsaleh et al., 2020). In addition to physiological parameters, monitoring environmental parameters can also play an essential role in transmitting some infectious diseases such as water-associated and vector-borne infectious diseases (Yang et al., 2012). The environmental parameters are also crucial for assessing their possible effects on emerging infectious diseases and supporting decision-making to control the disease effectively, as has been done by (Poirier et al., 2020) during the early phase of the ongoing COVID-19 pandemic.

We implemented an extension of the IoT-Stream ontology using Protégé ver. 5.5.0 (Musen & Team, 2015) and evaluated the IoT4PHM ontology using competency questions (CQ) listed in Table 1.

We performed semantic enrichment by mapping data to the ontology classes using the open-source tool called Karma, which is a data integration tool that allows the transformation of the data into Linked Data by creating URIs for entities (Knoblock et al., 2012).

Finally, we demonstrate basic query search on the enriched RDF using SPARQL Protocol and RDF Query Language (SPARQL) to extract some patient health information. We used a SPARQL processor called ARQ, which is part of the Apache Jena framework (The Apache Software Foundation, n.d.).

Table 1: Competency Questions.

| | |
|---|---|
| CQ1 | What types of patient data are collected? |
| CQ2 | Is there any information gathered that can be associated with a worse disease prognosis? |
| CQ3 | What are the possible types of quantity kind to monitor patient health in the context of an infectious disease outbreak? |
| CQ4 | What are the main types of patient management that can be utilized for patient triage according to the detected event? |

## 4 ILLUSTRATION OF IoT4PHM ONTOLOGY AND PROOF OF CONCEPT

We illustrated how the IoT4PHM ontology enables the semantic representation of the vital sign streams to monitor the patient's health condition. We present a proof-of-concept of ontology-based semantic modeling and semantic enrichment of vital sign data from the two different data sources.

First, we built the heart rate dataset's semantic model from the EMR vital sign dataset. We semi-automatically performed the semantic enrichment by mapping data to the ontology entities using the Karma Data Integration Tool. In addition, to handle the volume and velocity of streaming ECG data from the Radar Vital Sign dataset and realize its real-time semantic enrichment, we performed summarization to enrich ECG streams, inspired by the previous study (Pacha et al., 2020). To summarize the 10-minute ECG streams, we applied a sliding window to compute the average heart rate over a set of one-minute periods. After the summarization step, the same semantic model used to annotate vital signs from the EMR vital sign dataset can also be applied to automatically annotate the summarized ECG data streams from the Radar Vital Sign dataset. Figure 2 shows the enrichment of the first sliding window from the patient's ECG stream whose ID is "GDN0021" (from the Radar Vital Sign dataset) in a turtle format. Semantic IoT data can later be queried, interpreted, and reasoned to generate new information and knowledge (Zgheib et al., 2020).

After converting patient data into RDF using semantic enrichment, we performed a basic SPARQL query search to extract some patient's information.

Figure 3 shows the SPARQL query we run to identify the female patients older than or at the age of 70, with a respiration rate greater than or equal to 25 breaths per minute, and the search results.

## 5 CONCLUSIONS

There is a lack of an ontology that enables IoT-based patient health monitoring and management in the context of infectious disease outbreaks. To answer the research question, we extended the IoT-Stream ontology to create the IoT4PHM ontology to enable patient health monitoring and management for the prevention, early detection, and mitigation of patient deterioration during infectious disease outbreaks.

We evaluated the IoT4PHM ontology using four CQs in Table 1. The answer to the CQ1 is that the current version of the ontology can store the patient's ID, age, gender, symptoms, and (if any) recent contact with a wild animal. Such information is essential to collect, especially at an early phase of a newly emerging infectious disease outbreak, for investigating possible high-risk groups and symptoms and virus spillover from wildlife.

```
iot4phm:Patient iot4phm:hasID "GDN0021" ;
iot4phm:hasAge "24" ;
iot4phm:hasGender "M" ;
iot4phm:hasStreamObservation "GDN0021-meanHR-so1" .
iot-stream:StreamObservation1 rdf:type iot-
stream:StreamObservation ;
sosa:hasSimpleResult "80" ;
sosa:resultTime "2020-05-22 16:45:00" ;
iot-stream:belongsTo iot-stream:IoTStream ;
iot-stream:madeBySensor sosa:Sensor .
sosa:Sensor rdf:label "59230" ;
iot-lite:hasUnit "beats/minute" .
sosa:Sensor iot-lite:hasQuantityKind _:QuantityKind1 .
_:QuantityKind1 rdf:type qu:QuantityKind .
iot4phm:PhysiologicalParameter iot4phm:isSubQuantityKindof
_:QuantityKind1 ;
rdfs:label "Heart rate" .
iot-stream:IoTStream1 rdf:label "GDN0021-meanHR-01" ;
rdf:type iot-stream:IotStream ;
iot-stream:streamStart "2020-05-22 16:44:00" ;
iot-stream:streamEnd "2020-05-22 16:45:00" ;
iot-stream:generatedBy sosa:Sensor ;
qoi:hasQuality _:Quality1 .
_:Quality1 rdf:type qoi:Quality .
_:Timeliness1 rdf:subClassOf _:Quality1 ;
rdf:type qoi:Timeliness .
qoi:Frequency1 rdf:subClassOf _:Timeliness1 ;
qoi:hasAbsoluteValue "1 minute" ;
rdf:type qoi:Frequency .
iot-stream:IoTStream1 geo:location geo:Point .
geo:Point geo:lat "45,2977477" ;
geo:long "11,6583815" ;
iot-lite:altRelative "Vo Euganeo, Veneto, Italy, point" .
```

Figure 2: An excerpt of ontology-based semantic enrichment for the first sliding window of the ECG stream for patient "GDN0021".

```
SELECT ?hasID ?StreamObservation ?hasSimpleResult
WHERE
{
  ?Patient iot4phm:hasID ?hasID .
  ?Patient iot4phm:hasAge ?hasAge .
  ?Patient iot4phm:hasGender ?hasGender .
  ?Patient iot4phm:hasStreamObservation ?StreamObservation .
  ?StreamObservation sosa:hasSimpleResult ?hasSimpleResult .

  FILTER (?hasAge >= "70" && ?hasGender = "1" && ?hasSimpleResult != "NaN"
          && ?hasSimpleResult >= "25")
}
```

```
--------------------------------------------------------------------------------------------------
hasID     | StreamObservation                                                    | hasSimpleResult
==================================================================================================
"p019423" | <http://localhost:8080/source/StreamObservation8/p019423-RR-raw-so8> | "28.5"
"p002525" | <http://localhost:8080/source/StreamObservation6/p002525-RR-raw-so6> | "27"
"p004758" | <http://localhost:8080/source/StreamObservation2/p004758-RR-raw-so2> | "25"
"p018124" | <http://localhost:8080/source/StreamObservation9/p018124-RR-raw-so9> | "26"
"p018124" | <http://localhost:8080/source/StreamObservation3/p018124-RR-raw-so3> | "25"
"p018124" | <http://localhost:8080/source/StreamObservation6/p018124-RR-raw-so6> | "25"
"p018124" | <http://localhost:8080/source/StreamObservation8/p018124-RR-raw-so8> | "29"
"p018124" | <http://localhost:8080/source/StreamObservation7/p018124-RR-raw-so7> | "32.5"
"p018124" | <http://localhost:8080/source/StreamObservation4/p018124-RR-raw-so4> | "26"
--------------------------------------------------------------------------------------------------
```

Figure 3: An excerpt of the SPARQL query and results for extracting the IDs of the patients with a respiration rate ≥ 25.

The answer to the CQ2 is that the IoT4PHM ontology has the UnderlyingHealthCondition class that can capture risk factors associated with the increased risk of developing complications. This information is critical to identify high-risk groups who need to be prioritized for the treatment and public health interventions.

The answer to the CQ3 is that both physiological and environmental parameters are included as subclasses of the QuantiyKind class. Since environmental factors play a crucial role in transmitting some infectious diseases, they significantly impact public health strategies.

Finally, the answer to the CQ4 is the patient management can be classified as either HealthEducation, ReferralToHealthcare, and EmergencyAlert, depending on the severity of the detected event (e.g., an early warning score). The classification contributes to ensuring the limited resources are effectively allocated to those who need most and prevent overburdening healthcare systems.

Therefore, the IoT4PHM ontology successfully addressed all the CQs and can potentially

be used for effective patient health monitoring and management during infectious disease outbreaks.

In our future study, we will further extend the ontology by adding relevant concepts to annotate the aggregated individual patient data to obtain population-level data.

Furthermore, to evaluate the capability of the IoT4PHM ontology more rigorously and improve the quality of the ontology, we will invite domain experts to assess the ontology in terms of accuracy, clarity, and completeness. We also plan to evaluate the validity of our ontology through a series of real annotation scenarios.

## ACKNOWLEDGEMENTS

We sincerely appreciate Dr. Stefano Bonacina at Health Informatics Centre, Department of learning, informatics, management, and ethics, Karolinska Institutet, for his constructive advice on software usage and the ontology extension.

## REFERENCES

Abrams, E. M., & Szefler, S. J. (2020). COVID-19 and the impact of social determinants of health. *The Lancet Respiratory Medicine*, *8*(7), 659–661.

Alirezaie, M., Renoux, J., Köckemann, U., Kristoffersson, A., Karlsson, L., Blomqvist, E., Tsiftes, N., Voigt, T., & Loutfi, A. (2017). An Ontology-based Context-aware System for Smart Homes: E-care@home. *Sensors (Basel, Switzerland)*, *17*(7). https://doi.org/10.3390/s17071586

Bermudez-Edo, M., Elsaleh, T., Barnaghi, P., & Taylor, K. (2017). IoT-Lite: a lightweight semantic model for the internet of things and its use with dynamic semantics. *Personal and Ubiquitous Computing*, *21*(3), 475–487. https://doi.org/10.1007/s00779-017-1010-8

Biswanath, D. (2017). Examining the interrelatedness between ontologies and Linked Data. *Library Hi Tech*, *35*(2), 312–331. https://doi.org/10.1108/LHT-10-2016-0107

Carbonaro, A., Piccinini, F., & Reda, R. (2018). Integrating Heterogeneous Data of Healthcare Devices to enable Domain Data Management. *Journal of E-Learning and Knowledge Society*, *14*, 45–56. https://doi.org/10.20368/1971-8829/1450

Clark, A., Jit, M., Warren-Gash, C., Guthrie, B., Wang, H. H. X., Mercer, S. W., Sanderson, C., McKee, M., Troeger, C., Ong, K. L., Checchi, F., Perel, P., Joseph, S., Gibbs, H. P., Banerjee, A., Eggo, R. M., Nightingale, E. S., O'Reilly, K., Jombart, T., … Jarvis, C. I. (2020). Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. *The Lancet Global Health*, *8*(8), e1003–e1017. https://doi.org/10.1016/S2214-109X(20)30264-3

Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W. D., Le Phuoc, D., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., … Taylor, K. (2012). The SSN ontology of the W3C semantic sensor network incubator group. *Journal of Web Semantics*, *17*, 25–32. https://doi.org/https://doi.org/10.1016/j.websem.2012.05.003

Elsaleh, T., Enshaeifar, S., Rezvani, R., Acton, S. T., Janeiko, V., & Bermudez-Edo, M. (2020). IoT-Stream: A Lightweight Ontology for Internet of Things Data Streams and Its Use with Data Analytics and Event Detection Services. *Sensors (Basel, Switzerland)*, *20*(4), 953. https://doi.org/10.3390/s20040953

Epstein, R. M., Fiscella, K., Lesser, C. S., & Stange, K. C. (2010). Why the nation needs a policy push on patient-centered health care. *Health Affairs (Project Hope)*, *29*(8), 1489–1495. https://doi.org/10.1377/hlthaff.2009.0888

Fariha, M. A. Z., Ikeura, R., Hayakawa, S., & Tsutsumi, S. (2020). Analysis of Pan-Tompkins Algorithm Performance with Noisy ECG Signals. *Journal of Physics: Conference Series*, *1532*, 12022. https://doi.org/10.1088/1742-6596/1532/1/012022

Ganz, F., Barnaghi, P., & Carrez, F. (2016). Automated Semantic Knowledge Acquisition From Sensor Data. *IEEE Systems Journal*, *10*(3), 1214–1225. https://doi.org/10.1109/JSYST.2014.2345843

Golinelli, D., Boetto, E., Carullo, G., Nuzzolese, A. G., Landini, M. P., & Fantini, M. P. (2020). Adoption of

Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literature. *Journal of Medical Internet Research*, *22*(11), e22280. https://doi.org/10.2196/22280

Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. In *Synthesis lectures on the semantic web: theory and technology* (Vol. 1, Issue 1). Morgan & Claypool Publishers.

Jabbar, S., Ullah, F., Khalid, S., Khan, M., & Han, K. (2017). Semantic Interoperability in Heterogeneous IoT Infrastructure for Healthcare. *Wireless Communications and Mobile Computing*, *2017*, 9731806. https://doi.org/10.1155/2017/9731806

Janowicz, K., Haller, A., Cox, S. J. D., Le Phuoc, D., & Lefrançois, M. (2019). SOSA: A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics*, *56*, 1–10. https://doi.org/10.1016/j.websem.2018.06.003

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, *451*(7181), 990–993. https://doi.org/10.1038/nature06536

Kelly, J. T., Campbell, K. L., Gong, E., & Scuffham, P. (2020). The Internet of Things: Impact and Implications for Health Care Delivery. *Journal of Medical Internet Research*, *22*(11), e20135. https://doi.org/10.2196/20135

Knoblock, C. A., Szekely, P., Ambite, J. L., Goel, A., Gupta, S., Lerman, K., Muslea, M., Taheriyan, M., & Mallick, P. (2012). *Semi-automatically Mapping Structured Sources into the Semantic Web BT - The Semantic Web: Research and Applications* (E. Simperl, P. Cimiano, A. Polleres, O. Corcho, & V. Presutti (eds.); pp. 375–390). Springer Berlin Heidelberg.

Lim, S., & Rahmani, R. (2020). Toward Semantic IoT Load Inference Attention Management for Facilitating Healthcare and Public Health Collaboration: A Survey. *Procedia Computer Science*, *177*, 371–378. https://doi.org/10.1016/j.procs.2020.10.050

Musen, M. A., & Team, P. (2015). The Protégé Project: A Look Back and a Look Forward. *AI Matters*, *1*(4), 4–12. https://doi.org/10.1145/2757001.2757003

Nasajpour, M., Pouriyeh, S., Parizi, R. M., Dorodchi, M., Valero, M., & Arabnia, H. R. (2020). Internet of Things for Current COVID-19 and Future Pandemics: an Exploratory Study. *Journal of Healthcare Informatics Research*, *4*(4), 325–364. https://doi.org/10.1007/s41666-020-00080-6

Pacha, S., Murugan, S. R., & Sethukarasi, R. (2020). Semantic annotation of summarized sensor data stream for effective query processing. *The Journal of Supercomputing*, *76*(6), 4017–4039. https://doi.org/10.1007/s11227-017-2183-7

Perera, C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2014). Context Aware Computing for The Internet of Things: A Survey. *IEEE Communications Surveys & Tutorials*, *16*(1), 414–454. https://doi.org/10.1109/SURV.2013.042313.00197

Poirier, C., Luo, W., Majumder, M. S., Liu, D., Mandl, K. D., Mooring, T. A., & Santillana, M. (2020). The role of environmental factors on transmission rates of the COVID-19 outbreak: an initial assessment in two spatial scales. *Scientific Reports*, *10*(1), 1–11.

Reyna, M. A., Josef, C. S., Jeter, R., Shashikumar, S. P., Westover, M. B., Nemati, S., Clifford, G. D., & Sharma, A. (2020). Early Prediction of Sepsis From Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. *Critical Care Medicine*, *48*(2), 210–217. https://doi.org/10.1097/CCM.0000000000004145

Schellenberger, S., Shi, K., Steigleder, T., Malessa, A., Michler, F., Hameyer, L., Neumann, N., Lurz, F., Weigel, R., Ostgathe, C., & Koelpin, A. (2020). A dataset of clinically recorded radar vital signs with synchronised reference sensor signals. *Scientific Data*, *7*(1), 291. https://doi.org/10.1038/s41597-020-00629-5

Sedghamiz, H. (n.d.). *Complete Pan Tompkins Implementation ECG QRS detector*. MATLAB Central File Exchange.

Sheth, A., Henson, C., & Sahoo, S. (2008). Semantic Sensor Web. *Internet Computing, IEEE*, *12*, 78–83. https://doi.org/10.1109/MIC.2008.87

Strang, T., & Linnhoff-Popien, C. (2004). A Context Modeling Survey. In *Proceedings of the Workshop on Advanced Context Modeling, Reasoning and Management as Part of UbiComp*.

Swayamsiddha, S., & Mohanty, C. (2020). Application of cognitive Internet of Medical Things for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *14*(5), 911–915. https://doi.org/10.1016/j.dsx.2020.06.014

The Apache Software Foundation. (n.d.). *ARQ - A SPARQL Processor for Jena* (2.0). https://jena.apache.org/documentation/query/index.html

Xu, B., Gutierrez, B., Mekaru, S., Sewalk, K., Goodwin, L., Loskill, A., Cohn, E. L., Hswen, Y., Hill, S. C., Cobo, M. M., Zarebski, A. E., Li, S., Wu, C.-H., Hulland, E., Morgan, J. D., Wang, L., O'Brien, K., Scarpino, S. V., Brownstein, J. S., … Kraemer, M. U. G. (2020). Epidemiological data from the COVID-19 outbreak, real-time case information. *Scientific Data*, *7*(1), 106. https://doi.org/10.1038/s41597-020-0448-0

Yang, K., LeJeune, J., Alsdorf, D., Lu, B., Shum, C. K., & Liang, S. (2012). Global distribution of outbreaks of water-associated infectious diseases. *PLoS Negl Trop Dis*, *6*(2), e1483.

Zgheib, R., Kristiansen, S., Conchon, E., Plageman, T., Goebel, V., & Bastide, R. (2020). A scalable semantic framework for IoT healthcare applications. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-020-02136-2