Formal Ontology in Information Systems
S. Borgo et al. (Eds.)

doi:10.3233/978-1-61499-910-2-239

SAREF4health: IoT Standard-Based Ontology-Driven Healthcare Systems

João MOREIRA[a,1], Luís FERREIRA PIRES[a], Marten van SINDEREN[a], Laura DANIELE[b]
[a] *University of Twente, Enschede, The Netherlands*
[b] *TNO, Netherlands Organisation for Applied Scientific Research*

Abstract. Recently, a number of ontology-driven healthcare systems have been leveraged by the Internet-of-Things (IoT) technologies, which offer opportunities to improve patient monitoring and abnormal situation detection with support of medical wearables and cloud infrastructure. Usually, these systems rely on IoT ontologies to represent sensor data observations. The ETSI Smart Appliances REFerence (SAREF) IoT ontology is an extensible industry-oriented standard. In this paper, we discuss the verbosity problem of SAREF when used for real-time electrocardiography (ECG), emphasizing the requirement of representing time series. We compared the main ontologies in this context according to quality, message size (payload), IoT-orientation and standardization. We also introduce a SAREF4health extension to tackle the verbosity problem. In the SAREF4health development we followed ontology-driven conceptual modelling, in which an ECG ontology grounded in the Unified Foundational Ontology (UFO) plays the role of a reference model. The methodology was enhanced by a standardization procedure and considers the RDF serialization of the HL7 Fast Healthcare Interoperability Resources (FHIR) standard. The validation of SAREF4health includes the use cases of an early warning system that uses ECG data to detect accidents with truck drivers in a port area. A prototype that integrates an existing ECG wearable with cloud infrastructure demonstrates the performance impact of SAREF4health considering IoT constraints. Our results show that SAREF4health is adequate to enable semantic interoperability of IoT solutions that need to deal with frequency-based time series. Design decisions regarding the trade-off between ontology quality and aggregation representation are also discussed.

Keywords. ontology-driven healthcare system, Internet of Things, ontology-driven conceptual modelling, health ontology, SAREF. ECG monitoring

1. Introduction

In the past years, a number of ontology-driven e-health solutions, improved with IoT technologies, were proposed [2,9,15]. Current IoT electrocardiography (ECG or EKG) approaches use medical wearables and cloud infrastructure for real time monitoring of cardiac behaviour [1-3], which improves patient monitoring for automatic identification of abnormal situations.

Usually, these systems rely on IoT ontologies to represent sensor data observations. In this context, several ontology-based IoT approaches have been proposed [4, 5], from which the Smart Appliances REFerence (SAREF) ontology is an extensible IoT

[1] Corresponding Author.

reference ontology developed in close connection with the IoT industry [6]. SAREF is standardized by the European Telecommunications Standards Institute (ETSI) and includes extensions, e.g., smart energy, environment (SAREF4envi) and buildings [7], being considered by the European Commission as “a first ontology standard in the IoT ecosystem and sets a template and a base for development of similar standards for the other verticals to unlock the full potential of IoT” [7].

In this paper, we address the verbosity problem of SAREF messages in IoT scenarios of real-time electrocardiography (ECG), where data needs to be represented as frequency-based time series of measurements observed by sensors. We present a study that compares the main semantic models in this context, according to four characteristics: quality, message size (payload), IoT-orientation and standardization. To address the verbosity problem, we introduce SAREF4health, an extension of SAREF leveraged by ontology-driven conceptual modelling, in which an ECG ontology grounded in the Unified Foundational Ontology (UFO) plays the role of reference model. The approach combines the method of ontology-driven conceptual modelling with standardization initiatives, e.g., HL7 Fast Healthcare Interoperability Resources (FHIR), and best practices for RDF implementation of stream data.

In the first part of the SAREF4health validation, instances of SAREF4health were created and competency questions were answered with SPARQL queries. In the second part of the validation, we measured performance through a prototype that simulates use cases of an IoT early warning system (EWS) monitoring ECG data to detect accidents with truck drivers in a port area. Finally, we discuss the trade-offs regarding the impact of design decisions and ontology quality.

Section 2 presents some background on standardization initiatives for semantic interoperability in IoT domain, IoT solutions for ECG monitoring and healthcare semantic models. Section 3 describes the SAREF verbosity problem. Section 4 describes the SAREF4health development and validation. Section 5 summarizes our contributions, lessons learned and future work.

2. Background

2.1. Standards for IoT semantic interoperability

Among the numerous IoT ontologies proposed in the literature, the W3C Semantic Sensor Network (SSN/SOSA) and the ETSI SAREF are the most prominent approaches [4, 5], which were rigorously developed during many years by ontologists and domain experts, being applied in several IoT use cases and supported by standardization initiatives. SSN/SOSA and SAREF are aligned with each other [8]. In particular, SAREF is pointed by the EC as the main reference IoT ontology that provides an extensible characteristic based on an ETSI procedure to standardize extensions of SAREF for vertical markets [7]. The motivation behind SAREF was that the market would continue to be too fragmented and powerless without a (protocol-independent) semantic layer that enables interoperability among the various smart appliances from different manufacturers. To that end, SAREF was created with the intention to interconnect different platforms, supporting the data exchange with different protocols.

To avoid semantic interoperability issues when extending SAREF, an ETSI standard procedure should be followed whenever a SAREF extension is submitted for acceptance. This procedure is based on the following best practices: (1) the extension is designed according to clarity, coherence, extensibility, minimal encoding bias and minimal ontological commitment criteria; (2) relevant stakeholders in the domain of interest should be involved in its development process; (3) the group/community that creates the extension should be committed to contribute to its maintenance; (4) it should not add concepts that are already present in SAREF or other extensions; (5) it needs to be properly documented and published.

2.2. IoT solutions for electrocardiography (ECG)

Several solutions for ECG are leveraged by IoT technologies for real time monitoring of cardiac behaviour [1-3]. For example, modern ECG units can transmit the records of the electrical signals measured from the skin through wireless communications (e.g. Bluetooth), and are becoming smaller, more comfortable and energy-efficient. Real-time outpatient ECG monitoring is known as mobile cardiac telemetry [9] and can stream data continuously to caregivers, enabling the immediate alarm upon an abnormal situation detected, which can be implemented through an early warning system (EWS) [10].

We have been developing an EWS within the INTER-IoT-EWS project [11], which requires the detection of emergencies and risks of accidents with truck drivers wearing the Shimmer3 ECG unit[2] while they are transporting goods in a port area. The goal is to decrease the risk of fatal accidents at the port, improving health prevention and enabling quick reaction. Non-functional requirements include the semantic integration of IoT assets, e.g., devices, gateways, brokers and applications. Functional requirements include the real-time detection of arrhythmia (bradycardia and tachycardia) from ECG data provided by the ECG device and the detection of vehicle collision/impact by processing the cross-axial function ($x^2 + y^2 + z^2$) of instantaneous tri-axial acceleration data, similarly to Fall-MobileGuard [12].

2.3. Semantic health data

Open Biological and Biomedical Ontology (OBO) Foundry is a popular ontology-driven initiative in the biomedical informatics field, which provides a number of domain ontologies leveraged by the Basic Formal Ontology (BFO), a foundational ontology. For example, OBO provides the Eagle-I Resource Ontology, which includes several terms for health-related instruments and protocols, such as ECG monitoring device[3]. SNOMED CT[4] and LOINC[5] are other relevant ontology-based approaches for clinical health terminologies, since they define terms like ECG voltage and ECG abnormal finding.

A well-founded ECG ontology (UFO ECG) [13] was designed with support of an ontological analysis of existing health standards and ontology-driven conceptual modelling with UFO. The main goal of UFO ECG is to serve as a reference "unified Electronic Health Record (EHR) model", providing mappings to the three most common standards that support the representation of ECG data, including the HL7

[2] http://www.shimmersensing.com/products/ecg-development-kit
[3] http://purl.obolibrary.org/obo/ERO_0000835
[4] http://bioportal.bioontology.org/ontologies/SNOMEDCT
[5] http://bioportal.bioontology.org/ontologies/LOINC

annotated ECG standard (HL7 aECG) [14]. The use of HL7 v3 along with IEEE1451 in embedded devices to achieve end to end semantic interoperability on health systems is exploited in [15]. The HL7 aECG [16] is one of these standards, which was chosen by the US Food and Drug Administration (FDA) for clinical trials, implemented as a common lexicon approach, i.e., using XML schemas, running nowadays in several hospital information systems. A review of ECG storage formats can be found in [17], which also includes SCP-ECG and DICOM. This review concludes that aECG inherits the verbosity of XML, since it produces large messages (25 times larger than a compressed approach), so that it may only be used in clinical drug trials and presents conceptual design issues that hinder its instantiation.

HL7 Fast Healthcare Interoperability Resources (FHIR) is an emerging open standard that is considered as the evolution of the HL7 standards for the new generation of health data exchange. HL7 FHIR defines a collection of "resources" (data model elements) that can be mixed and adapted for particular clinical contexts through the "profiling" process. Along with W3C semantic web community, HL7 developed an RDF representation of the FHIR data model, which can be described and validated with Shape Expressions (ShEx) [18].

The central FHIR element is the *Observation* resource[6], which is used to support diagnosis and monitor progress, and is recommended for monitoring vital signs and device measurements. In particular, the "EKG example using Sampled Data" [7] demonstrates how to serialize ECG data using the *Observation* resource, which offers the property *valueSampledData*, responsible to represent the ECG sample sequence as a series of measurements of the heart electrical activity through the *SampledData* element, similarly to HL7 aECG. A *SampledData* provides a concise way to handle the data produced by devices that sample a particular physical state at a high frequency. A typical use for this is to represent the output of an ECG device. The data type includes a series of raw decimal values, which are mostly simple integers, along with adjustments for scale and factor.

3. Problem definition

Allowing the exchange of lightweight messages among medical wearables, gateways and cloud infrastructure is an important requirement of IoT solutions when processing big data. One of the main problems is that the verbosity of messages impacts the data exchange performance and cloud infrastructure costs. For example, the costs of the Microsoft Azure cloud gateway (IoT Hub) varies according to the message payload size[8]. A common approach in IoT to tackle this problem is to aggregate measurement data at the gateway level according to a certain frequency, transmitting time series from the local gateway to the cloud from time to time [19]. The drawback of this approach is that when aggregating a series of measurements of one element, metadata about each specific measurement is lost, affecting the ontological expressiveness of the messages.

UFO ECG represents series of measurements with the *Sample sequence* element, equivalent to FHIR element *Sampled Data*. According to UFO ECG, *Observation*

[6] www.hl7.org/fhir/observation.html

[7] www.hl7.org/fhir/observation-example-sample-data.json.html

[8] docs.microsoft.com/en-us/azure/iot-hub/iot-hub-devguide-messages-construct

series represents the complex event composed by (a *part-of* relationship) *Observations* that are evenly spaced in time, which are carried out in a *Recording session*. A *Sample sequence* is a collective, i.e., an ultimate sortal, which has parts that play the same functional role in the whole, represented as an ordered sequence of data units that results from an *Observation series*. OpenEHR emphasizes this requirement of time series representation and includes the concept of "a history of events, i.e., as a time series, allowing all software to access data in a uniform way". The root object (*History* class) provides a list of events as attribute.

SAREF does not define a term for the time series concept, only allowing the representation of each granular measurement of a time series through the *Measurement* element. Although this approach improves the expressiveness of the data representation, thus the ontology quality, it produces an overhead to the message size when serialized. Therefore, a verbose message is generated when instantiating ECG data with SAREF, compared to the very same data represented with FHIR, UFO ECG or OpenEHR. For example, our experiments on ECG data serialization, when comparing a JSON-LD message that instantiates SAREF individual *Measurements* against the similar message in FHIR RDF, but with *SampledData* element, we found out that the SAREF message size is fifty times bigger (more verbose) than the FHIR message. This was calculated based on our performance evaluation [20], where an ECG device configured for 256Hz frequency provide data to a semantic gateway, which accumulates data each 5 seconds before sending the message to the cloud, thus 1280 (5 x 256) measurements per message. While the equivalent SAREF message size has around 5Mb, the FHIR message size has around 100Kb. The discrepancy between SAREF and FHIR messages is non-linear, i.e., the difference of message size grows exponentially according to the number of measurements.

While in FHIR (or UFO ECG or OpenEHR) a measurement is only a number added to the *data* element (property of *SampledData*) representing data of the time series, in SAREF a *Measurement* is a data structure that implements a number of properties (e.g. *isMeasuredIn* and *relatesToProperty*). Similarly to SAREF, W3C SSN/SOSA does this through the *Observation* class, alleviating this issue by introducing the *hasSimpleResult* property, which enables the direct link to a number (literal), but still requires the representation of other properties (e.g. *observedProperty*). Therefore, due to this verbosity issue, we conclude that default SAREF is not suitable for exchanging IoT-based ECG time series data.

Table 1. Comparison of semantic models for IoT-based ECG monitoring systems

Characteristic	UFO ECG	FHIR RDF	SAREF	SSN/SOSA
Quality	+	-	+	+
Message size (payload)	+	+	-	-
IoT-oriented	-	+-	+	+
Standardized	-	HL7	ETSI	W3C

We studied alternatives, identifying their benefits and drawbacks. Table 1 compares the semantic models[9] considered in this study. The quality of the semantic models is the first characteristic we compared, following the best practices of ontology engineering described in [21], more specifically the best practices #8 and #12, which regard common ontology pitfalls and ontology reuse, respectively. While UFO ECG,

[9] We avoid the term "ontology" because we understand that an ontology describes the common sense knowledge domain rather than a simple RDF serialization.

SAREF and SSN/SOSA are built upon common conceptualization mechanisms, merging different concepts in the same class/property and reusing existing ontologies, FHIR RDF presents poor quality because it is a (semi) automatic serialization of FHIR lexicon standard. Thus, FHIR RDF is a straightforward serialization of the FHIR standard data model, which causes an overload of object properties, e.g., it has more than ten properties that represent the "description" concept, and lacks reuse of common terms, e.g., instead of having more than ten "description" terms, it could reuse *dc:description*.

The message size (payload) indicates whether the semantic model allows the representation of time series, as in UFO ECG and FHIR RDF. We also evaluated whether the semantic model is appropriate to the IoT context, i.e., whether it provides grammatical constructs for the main concepts of the IoT domain, e.g., sensor, actuator, protocol, observation, unit of measurement and measurement properties. Specific characteristics of the ECG domain were considered in this analysis, such as the mereology of an ECG device, the elements participating in an ECG recording session, the ECG leads responsible for measuring frequency-based sample sequences (time series). Since our goal is the development of an EWS for the detection of risks of truck accidents, we also considered how to represent and retrieve multiple sensors data, such as acceleration data to calculate tri-axial function (impact) and ECG data to calculate ECG waveform features. SSN/SOSA and SAREF are the most appropriate IoT reference ontologies, as mentioned earlier, while some FHIR RDF resources can be used to represent some of the IoT constructs. Finally, we evaluated whether the semantic model is standardized by a standardization body, such as HL7, ETSI and W3C for FHIR, SAREF and SSN/SOSA (respectively).

In our comparison study, we concluded that a standardized IoT ontology that provides adequate balance of quality and payload representing time series for ECG data is actually lacking. Therefore, the problem addressed in this paper is how to achieve these four characteristics in a single ontology, which led us to the development of SAREF4health as a SAREF extension.

Table 2. Competency questions to be responded by SAREF4health

ID	Textual description
CQ01	What is an ECG device and how it is composed (mereology)?
CQ02	What are the elements participating in an ECG recording session?
CQ03	What is an ECG lead, what are the types of ECG leads, what type of property an ECG lead measures and what type of measurement an ECG lead can measure?
CQ04	What is an ECG sample sequence?
CQ05	What is a time series of measurements?
CQ06	What is frequency (rate)measurement of an ECG sample sequence?
CQ07	How to represent tri-axial acceleration data from accelerometers of an ECG device?
CQ08	How to integrate measurements from multiple sensors (e.g., ECG leads, accelerometer and battery monitor) of an ECG device for near real-time (frequency-based) monitoring?

The main information requirements of this ontology regard the ECG and the IoT domains, described as competency questions (Table 2). While UFO ECG provides a high quality and deeper ontological analysis of the ECG domain, with rich descriptions following foundational categories, SAREF can be used as IoT reference model. Therefore, a set of competency questions was defined according to the main elements of these reference ontologies. The scope of these questions was narrowed with the information requirements of the use case validation of an EWS to detect truck accidents mentioned in Section 2.2.

4. SAREF4health

4.1. Methodology

We have been working on the "SEmantic Model-driven development for IoT Interoperability of emergenCy serviceS" (SEMIoTICS) framework [22], which is a set of technologies and guidelines for the development of semantic IoT EWSs. SAREF4health is part of the INTER-IoT-EWS solution, playing the role of semantic model for ECG data used in our IoT health platform [11]. Ontology-driven conceptual modelling is an important component adopted by the SEMIoTICS framework, used for the design of the EWS with support of an ontological language that provides structural and temporal predicates, as OntoUML [23]. In order to create the SAREF4health extension, we adopted the cyclical semantic enhancement process of UFO research, increasing the quality of the core ontology through model assessment based on formal lightweight verification and validation activities [24]. The UFO ECG ontology played the role of a unified well-founded EHR reference model representing the ECG domain. A side effect of this approach is that UFO ECG provides links to HL7 aECG, which FHIR is based upon, thus, enabling straightforward alignments to FHIR RDF.

The SAREF4health design was leveraged by the same interactive and iterative approach experienced in the other standardized extensions of SAREF, which was created in a transparent manner to allow stakeholders to provide input and follow the evolution of the work. The first step comprised the requirement collection to guide the implementation and validation of the ontology. Numerous information sources were analysed, as specifications, datasets, standards, APIs and data formats, as well as domain expert opinions and existing initiatives in the healthcare domain. The second step comprised the use case collection, specified in natural language. The third step comprised the purpose and scope definition of the ontology for the specific use cases regarding the monitoring of the cardiac behaviour of drivers in a way that abnormal situations, e.g., arrhythmia (bradycardia and tachycardia), could be detected by the emergency system, reflected in the competency questions.

4.2. Representing ECG time series

The main definitions used to support the SAREF4health elements are described in Table 3. Although UFO ECG and FHIR standard were carefully designed with the support of healthcare experts, we decided to review the terminology used by them. For example, why UFO ECG uses the term *Sample Sequence* while FHIR uses *Sampled Data* to describe the very same concept? Either way, the terms reflect the *Time Series* collective concept, i.e., a sequence of data units in successive equally spaced points in time: each sampled data unit plays the same role in the series (whole).

In SAREF4health, we termed the element *Time Series Measurements* since it refers to a time series of a sequence of measurements made by a device and it is a term often used in the measurement science (metrology)[10]. We avoided prefixing this term with *ECG* because it can be applied to other types of measurements. An *ECG Sample Sequence* is a *Time Series Measurements* that is measured in *Electric Potential* units (an array) and relates to the *Heart Electrical Activity* property. A *Sample* (UFO ECG) can be interpreted as a *Measurement* (SAREF), so we classify *Measurement* as a kind.

[10] https://plato.stanford.edu/entries/measurement-science/

A crucial design decision was to classify *Time Series Measurements* as *Measurement* in SAREF, for two reasons: (i) this representation adheres to the definition of *Measurement*, i.e., measured value (*Electric Potential* units) of a property (*Heart Electrical Activity*); (ii) we reused the same structure of SAREF regarding class axioms of object properties, e.g., *hasTimestamp*, *isMeasuredIn* and *relatesToProperty*. The main implications of this choice are twofold: (1) ontologically this specialization is incorrect, since a collective cannot specialize a kind; (2) the *hasValue* property limits the value domain of a *Measurement* to exactly one float number. The *hasValues* property was added to overcome this issue, in which a *Time Series Measurements* can instantiate this property multiple times as an array of float numbers. In the conceptual model, instead of adding this specialization, the relation *makesMeasurement* between an *ECG Lead* and an *ECG Sample Sequence* makes this design decision explicit.

Table 3. Definitions related to time series in UFO ECG and standards

Term	Source(s)	Textual definition
Sample sequence	UFO ECG	Collective: "ordered sequence of samples resulting from an *Observation series*" (ecgOnto:095).
Observation series	UFO ECG	Complex event: "Series of observations evenly spaced in time carried out in an ECG Recording session" (ecgOnto:093).
Sampled data Observation.component.valueSampledData	HL7 FHIR	"Data that comes from a series of measurements taken by a device, which may have upper and lower limits".
Time Series Observation	O&M (ISO 19156)	"observation whose result is a time-series".
Series	HL7 aECG	"Contains one or more sequence sets sharing a common frame of reference".
Series General Series Module	DICOM	A property of General ECG that "specifies the attributes that identify and describe general information about the *Series* within a *Study*". A *Series* is as a sequence of data sharing a common frame of reference.
History	OpenEHR	"Root object of a linear history, i.e. time series structure. For a periodic series of events, period will be set, and the time of each Event in the History must correspond"
Event	OpenEHR	"Defines the abstract notion of a single event in a series. This class is generic, allowing types to be generated which are locked to particular spatial types"
Recording device	UFO ECG	Kind: "Device used to acquire (to record) an ECG from a given Patient by means of electrodes. Also called electrocardiograph" (ecgOnto:087).
Recording device as recorder	UFO ECG	Role: "Recording device as it plays the role of an ECG recorder" (ecgOnto:088).
Recording session	UFO ECG	Complex event: "Medical service in which the Patient is subject of ECG recording by some *Recording device*. The *Recording session* (event) can be said to temporally coincide, albeit in a different level of abstraction, with the Observation series (event). In other words, these two events have the same time boundaries".
Lead	UFO ECG	Kind: "Viewpoint of the heart activity that emerges from an Observation series of the p.d. between two electrode placements on specific regions of the surface of the patient's body" (ecgOnto:096)
Lead	HL7 aECG	"A vector along which the heart's electrical activity is recorded as a waveform"
Patient	UFO ECG	Role: "Person as he/she plays the role of being subject of care, i.e., scheduled to receive, receiving, or having received a healthcare service (based on ISO/TC 18308:2003)"

Finally, following the same approach of UFO ECG with *Sample Rate* data type and FHIR with *period* data property of *SampledData*, a class axiom was added in *Time*

Series Measurements to relate it to a frequency (rate). For the definition of frequency, we reused SAREF4envi, importing the object property *has Frequency Measurement*.

An *ECG device* is usually referred as an ECG unit that plays the role of a *recorder* in the complex event (action) of *ECG Recording Session*. In SAREF, we can classify this complex action as a *Task* that an ECG device *accomplishes*. Therefore, *accomplishes* plays the role of the inverse of the *hasParticipant* relationship (same as *isAccomplishedBy*). The *hasParticipant* relationship between *ECG Recording Session* and *Person Under ECG Monitoring* is implemented with *dc:author* (Dublin Core).

In UFO, the *event* stereotype provides the relations *start* and *end* to limit the event temporal boundaries, as used in UFO ECG. SAREF provides the *hasTime* property, a "relationship to associate time information to an entity", thus, we specialized this relation as *hasStart* and *hasEnd*, adding them as class axioms of *Task*. In UFO ECG, the other participant in this event is a *Patient*, but in SAREF4health we adopted a different ontological commitment. We argue that someone does not need to be a patient (person under medical treatment) to participate in an ECG recording session. Thus, we introduced the *Person Under ECG Monitoring* element, i.e., a role of a *Living Person*, which is a phase exploited several times in the UFO research [25]. This ontological commitment is motivated by the EWS use case, since a driver is not necessarily a patient. In SAREF4health, a *Living Person* is a *Person*, which is an element imported from DUL and is aligned to schema.org. We only consider ECG specific to humans.

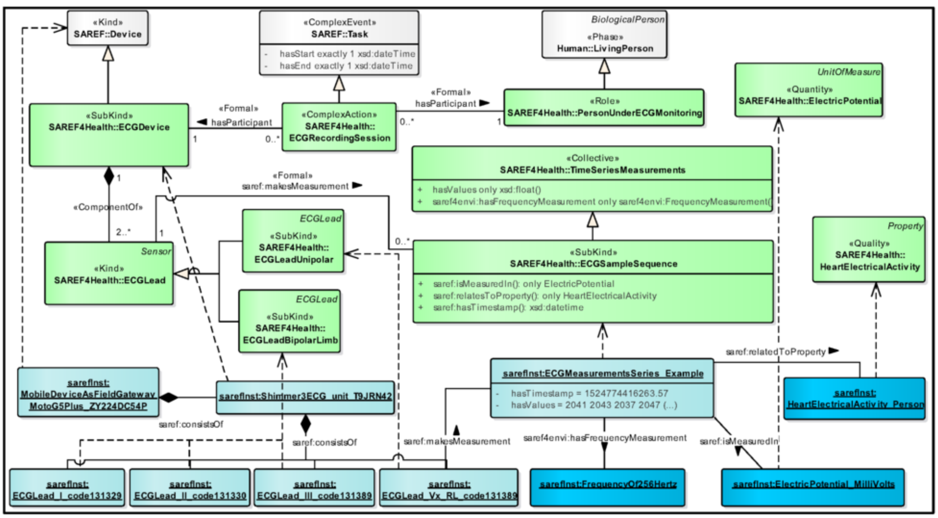

Figure 1. SAREF4health main elements (green) with instantiated examples (blue).

Usually, the frequency of an *ECG device* can be set through an API. This reflects the frequency of each *ECG Sample Sequence* measured during a *Recording Session*. Therefore, we added a class axiom to the *ECG device* element, the *has Frequency Measurement* property with range of only *Frequency Measurements* (from SAREF4envi). Although it seems redundant, this approach is required to differentiate the current frequency of a device from a frequency used in prior sample sequences. For example, the device Shimmer3 ECG can be set for sampling frequency (rate) of 512Hz, which is recommended for clinical grade ECG, i.e., 512 data samples per second or an interval of 0.002 second between two consecutive data samples, from 0.05Hz to 8000Hz range. After collecting some sample sequences and before sending the message to the gateway, the frequency is set to 256Hz and new sample sequences are collected. With our approach the message describes the current frequency of the device (256Hz) and the frequencies used in each collected sample sequence (512Hz and 256Hz).

An *ECG Device* registers the *Heart Electrical Activity* with support of electrodes attached to different places of the body (assuming that the heart is beating inside the body of a living person). Two electrodes enable the measuring of an *ECG lead*, which is an electrical vector characterized by the depolarization of the heart resulted by the electrical signal between the atria and the ventricles. Manufacturers commonly characterize an *ECG device* by the number of *ECG leads* it has. An ECG device is composed by extremity electrodes, which must be attached close to the left arm (LA), right arm (RA), left leg (LL) and the right leg (RL); and chest (precordial) electrodes, which can vary from one unit to six units (V_{1-6}). As a convention, lead I measures the electrical activity from RA to LA, lead II measures of the electrical activity from RA to LL, lead III measures the electrical activity from LA to LL. The rule lead I + lead III = lead II enables to derive a lead based on the other two. Leads I, II, III are known as *Bipolar Limb*. *Unipolar* leads measure the electrical activity from the Wilson's central terminal (negative pole) to each of the chest electrodes (positive poles). For example, the Shimmer3 ECG is a four-lead ECG device wired with four extremity electrodes and one chest electrode, enabling the measurement of three bipolar and one unipolar leads.

For the sake of simplicity and to avoid verbosity and follow the industry terminology, we decided to represent an *ECG device* according to its *ECG leads*, classifying an *ECG lead* as a *Sensor*, since a sensor "detects and responds to events or changes in the physical environment" (SAREF). A lead can be either bipolar or unipolar, and an *ECG device consists of*, at least, one *Bipolar Limb* and one *Unipolar* lead. An *ECG lead* measures the *Heart Electrical Activity* property and *makes measurements* of *ECG Sample Sequence*. Figure 1 represents these main elements in an OntoUML model.

4.3. Validation and discussion

The first part of the SAREF4health validation was performed by implementing the ontology as RDF, which was supported by the Menthor [24] and Protegé tools. Instances of SAREF4health were created (as examples), and basic SPARQL queries were executed to answer each competency question. As illustrated in some SAREF4health instances in Figure 1, the examples are based on the Shimmer3 ECG device, the device API (TinyOS), the mobile app (Android), which plays the role of field gateway, and an IoT platform context broker, playing the role of cloud gateway. Table 4 summarizes the results. Because of space limitation, we removed textual properties (e.g., labels and comments) from the results. The SPARQL queries used are based on the following template:

```
SELECT * WHERE { {?s1 ?p1 [Name of the term]} UNION {[Name of the term] ?p2 ?o2} . }
```

All competency questions could be answered, assessing the completeness of SAREF4health and, therefore, its semantic validity. Our ontological commitments were based on the definitions listed in Table 3 and were assessed by analysing the SPARQL results. For example, *time series* element (CQ05) is equivalent to *sample sequence* (UFO ECG) and *Sampled Data* (FHIR). The SPARQL result shows that the structure allows the representation of an ordered frequency-based sequence of float data. The semantic validity provided a basic quality level for SAREF4health, which was leveraged by the application of our methodology based on the best practices for ontology engineering. We argue that SAREF4health has a high quality because of the ontological foundations and other practical matters, such as making SAREF4health as dereferenceable as possible and checking and reusing popular and common ontologies. A limitation of SAREF4health is that it (still) lacks a concept to represent the physician(s) administering an ECG recording session, which is common in other standards (e.g., DICOM). To overcome this issue we suggest to add a class axiom to *ECG Recording Session* with the *dc:creator*, similarly to the approach with *dc:author*.

Table 4. Responding the competency questions with SPARQL queries

ID	Results	Description
CQ01	> rdfs:subClassOf saref:consistsOf min 1 ECGLeadUnipolar > rdfs:subClassOf saref:consistsOf min 1 ECGLeadBipolarLimb > rdfs:subClassOf saref:accomplishes only ECGRecordingSession > rdfs:subClassOf saref4envi:hasFrequencyMeasurement only saref4envi:FrequencyMeasurement > rdfs:subClassOf saref:Device < sarefInst:Shimmer3ECG_unit_T9JRN42 rdf:type	An ECG device is a device composed by at least one unipolar and one bipolar limb leads. It accomplishes the task of ECG recording session and has a specific frequency measurement.
CQ02	> rdfs:subClassOf dc:author only LivingPerson > rdfs:subClassOf saref:Task < ECGDevice saref:accomplishes only ECGRecordingSession < sarefInst:RecordingECGSession_01 rdf:type	An ECG recording session is a task in which a living person participates and is accomplished by an ECG device.
CQ03	> rdfs:subClassOf saref:measuresProperty only HeartElectricalActivity > rdfs:subClassOf saref:makesMeasurement only ECGSampleSequence > rdfs:subClassOf saref:Sensor < ECGLeadBipolarLimb rdfs:subClassOf < ECGLeadUnipolar rdfs:subClassOf	An ECG lead is a sensor that can be either bipolar limb or unipolar, which is able to measure a heart electrical activity, making measurements of the type ECG sample sequence.
CQ04	> rdfs:subClassOf saref:relatesToProperty only HeartElectricalActivity > rdfs:subClassOf saref:isMeasuredIn only ElectricPotential > rdfs:subClassOf TimeSeriesMeasurements < ECGLead saref:makesMeasurement only ECGSampleSequence < sarefInst:ECGMeasurementsSeries_Example001 rdf:type	An ECG sample sequence is a measurement time series that relates to the property of heart electrical activity, measures in an electric potential unit. ECG sample sequences are measured by ECG leads.
CQ05	> rdfs:subClassOf saref4envi:hasFrequencyMeasurement only saref4envi:FrequencyMeasurement > rdfs:subClassOf hasValues only xsd:float > rdfs:subClassOf saref:Measurement < ECGSampleSequence rdfs:subClassOf	A measurement time series is a measurement that has a frequency and a set of float values. Currently, the only type of time series measurements available is the ECG sample sequence.
CQ06	> saref4envi:FrequencyMeasurement > rdfs:subClassOf saref:Measurement > rdfs:subClassOf saref:isMeasuredIn exactly 1 saref4envi:FrequencyUnit > rdfs:subClassOf saref:relatesToProperty value saref4envi:Frequency	The frequency measurement type imported from SAREF4envi is a measurement which is measured in a frequency unit and relates to the frequency property.
CQ07	> sarefInst:Shimmer3ECG_unit_T9JRN42 rdf:type :ECGDevice ; saref:consistsOf sarefInst:AcceleroemeterX_ECGDevice ; saref:consistsOf sarefInst:AcceleroemeterY_ECGDevice ; saref:consistsOf sarefInst:AcceleroemeterZ_ECGDevice ; > sarefInst:AcceleroemeterX_ECGDevice rdf:type saref:Sensor ; saref:makesMeasurement sarefInst:Measurement_AccelerationX_001 ; saref:measuresProperty dim:Acceleration ; > sarefInst:Measurement_AccelerationX_001 rdf:type saref:Measurement ; saref:hasTimestamp "2018-04-22T22:15:30"^^xsd:dateTime ; saref:hasValue "100"^^xsd:float ; saref:isMeasuredIn unit:metrePerSecondSquared ; saref:relatesToProperty quantity:acceleration ;	Tri-axial acceleration data are represented according to each accelerometer sensor (X,Y,Z) of the ECG device. Each accelerometer sensor measures the property acceleration and makes acceleration measurements. An acceleration measurement is measured in metre per second squared, has a value and a timestamp.
CQ08	> sarefInst:Shimmer3ECG_unit_T9JRN42 rdf:type :ECGDevice ; saref4envi:hasFrequencyMeasurement sarefInst:FrequencyOf256Hertz ; saref:accomplishes sarefInst:RecordingECGSession_01 ; saref:consistsOf sarefInst:AcceleroemeterX_ECGDevice ; saref:consistsOf sarefInst:AcceleroemeterY_ECGDevice ; saref:consistsOf sarefInst:AcceleroemeterZ_ECGDevice ; saref:consistsOf sarefInst:ECGLead_III_code131389 ; saref:consistsOf sarefInst:ECGLead_II_code131330 ; saref:consistsOf sarefInst:ECGLead_I_code131329 ; saref:consistsOf sarefInst:ECGLead_Vx_RL_code131389 ; saref:consistsOf sarefInst:Shimmer3BatteryLevelSensor_T9JRN42 ; saref:hasManufacturer "Shimmer" ; saref:hasTypicalConsumption sarefInst:Shimmer3ECGBattery ;	The ECG device mereology is responsible for representing the device sensors through the *consists of* property. This structure allows (near) real-time monitoring by accumulating the measurements (made by the sensors). This includes the time series measurements (e.g. made by ECG leads) and isolated measurements (e.g. acceleration and battery level).

The second phase of the validation focused on providing evidence about the semantic interoperability improvement of an ontology-driven health IoT platform using SAREF4health in a common IoT scenario: connecting the device (Shimmer3 ECG unit) to a field gateway (Android smartphone) to a cloud gateway (MS Azure IoT Hub). A prototype was developed to improve the manufacturer's Android app (Shimmer Xamarin Capture) by translating the data received from the ECG device to SAREF4health and enabling different rates of data exchange between device-mobile and mobile-cloud, which is a common network bandwidth optimization requirement. This scenario also allowed us to exploit the SAREF4health approach for measurement time series. Therefore, the Android app plays the role of a semantic gateway, which serializes the SAREF4health messages with JSON-LD. Code Snippet 1 shows the method responsible for generating the ECG device messages in JSON-LD according to SAREF4health that are published in the cloud gateway.

The prototype is under functional tests of pre-defined accident simulations with truck drivers transiting in a port area, and the first results show that SAREF4health is an adequate data representation considering performance issues of IoT solutions. In particular, the message size (payload) was measured and indicates similar size to equivalent FHIR messages when representing time series. Therefore, SAREF4health can be classified as the most appropriate ontology for IoT-based ECG monitoring systems, according to the characteristics listed in Table 1.

Code snippet 1. Ontology-driven health IoT platform implementing data exchange of an ECG device

```
public JObject GetECGDeviceJSON_SAREF4health(List<JObject> listDevicesOfDevice, JObject recordingECGSession)
{
    string deviceId = "sarefInst:Shimmer3ECG_unit_T9JRN42_DeviceId";
    JObject eCGDeviceJSON = JObject.FromObject(new
    {
        comment = "Shimmer3 ECG unit (T9J-RN42): INTER-IoT-EWS project",
        label = "Shimmer3 ECG unit T9J-RN42",
        seeAlso = "http://www.shimmersensing.com/products/ecg-development-kit#specifications-tab"
    });
    eCGDeviceJSON.Add("@id", deviceId);
    eCGDeviceJSON.Add("@type", JArray.Parse("[ 'saref4health:ECGDevice', 'saref:Device']"));
    eCGDeviceJSON.Add("saref4envi:hasFrequencyMeasurement", hasFrequencyMeasurement);
    eCGDeviceJSON.Add("saref:accomplishes", recordingECGSession);
    eCGDeviceJSON.Add("saref:hasManufacturer", "Shimmer");
    eCGDeviceJSON.Add("saref:hasTypicalConsumption", hasTypicalConsumption);
    eCGDeviceJSON.Add("saref:consistsOf", JToken.FromObject(listDevicesOfDevice));

    return eCGDeviceJSON;
}
```

Currently, SAREF4health is not standardized yet, some activities of the ETSI procedure (described in Section 2.1) need to be addressed. At first, a specialists' task force needs to be created to enable interactive validation with domain experts and other stakeholders, creating new validation scenarios for ECG monitoring. Furthermore, the scope of the current version of SAREF4health may need to be extended to include new healthcare use cases besides ECG monitoring, which will require a new ontology engineering cycle. A plan for SAREF4health maintenance is under construction.

Besides the SAREF4health ontology, an important contribution of this paper is the methodology dealing with both conceptual modelling and implementation concerns, where design decisions must be clearly exposed to understand the discrepancies between the conceptual model and the final ontology. We argue that SAREF4health has higher ontological expressiveness because of its completeness for the ECG domain, since "the notion of completeness at the level of individual specifications is related to the notion of ontological expressiveness" [25]. Moreover, we believe that construct redundancy [26] was eliminated from SAREF4health regarding the elements about

Observation from UFO ECG. Although these terms are ontologically correct, in UFO ECG they bring additional complexity, represented implicitly in SAREF through the *make measurement* property and the *Measurement* class. While in SAREF a device makes a measurement, which has a timestamp, a value and is measured in a specific unit; in UFO ECG a device carries an observation, which has a timestamp and this observation produces a sample, which has a value and is measured in a specific unit.

Finally, as a side effect of this work, clear mappings between UFO ECG to FHIR were identified, as (UFO) *Observation series* to (FHIR) *Observation*, (UFO) *Sample sequence to* (FHIR) *Sampled Data*, (UFO) *Sample sequence > sample-sequence-of* is the inverse of *(FHIR) Observation > valueSampledData*. Other mappings can be easily extracted from the results of our ontological analysis described in Section 4.2. The SAREF4health ontology (TTL) and the prototype are available for download[11].

5. Conclusion

Although ETSI SAREF and W3C SSN/SOSA are standardized IoT reference ontologies, they only provide a verbose way to describe time series data in common healthcare scenarios, as for ECG monitoring. In this paper, we describe an extension of SAREF that addresses this problem, combining ontology-driven conceptual modelling, standardization initiatives and RDF implementation of stream data. SAREF4health is able to represent real-time ECG time series of sensor measurements that are exchanged between the field (mobile device) and the cloud (context broker) gateways.

Our validation showed that a trade-off between ontology quality and lightweight data serialization was a crucial aspect on the design of the SAREF4health ontology. The use of an ECG reference ontology grounded in UFO theory played a major role to improve the ontology quality, resulting on the improvement of the semantics of the RDF implementation. Furthermore, the reuse of standardized ontologies showed to be essential to understand both IoT and healthcare domains, as well as implementation constraints. SAREF4health addressed the performance requirement regarding the impact of time series measurements in the message size (payload), i.e., the verbosity problem, validated through a prototype of an ontology-driven health IoT platform, which transmits data from the device to a mobile to the cloud in different frequencies.

Future work includes a formal empirical validation to compare SAREF4health with other ontologies studied here, especially HL7 FHIR RDF. Semantic translations between SAREF4health and FHIR RDF are under development to address the requirements of integrating different IoT platforms that support data acquisition for an emergency system. Semantic loss and processing impact of these translations will be measured. The emergency system will be tested through real scenarios of accident prevention and response in the port of Valencia.

Finally, after this full validation, SAREF4health will be submitted to the ETSI standardization process. Independent of whether SAREF4health will ever become a standard, we believe that this work provides an important contribution towards a higher-quality FHIR ontology, rather than a simple RDF representation of the FHIR lexicon data model. Although HL7 FHIR lexicon standard is widely adopted and used in industry (e.g., by Google and Apple), its semantic model is still in early stages.

---

[11] https://github.com/jonimoreira/INTER-IoT-EWS

Acknowledgments Work carried out under the CAPES funding BEX 1046/14-4.

References

1. Xia, H., I. Asif, and X. Zhao, *Cloud-ECG for real time ECG monitoring and analysis.* Computer Methods and Programs in Biomedicine, 2013. **110**(3): p. 253-259.
2. Rahmani, A.M., et al., *Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: A fog computing approach.* Future Generation Computer Systems, 2018. **78**: p. 641-658.
3. Hossain, M.S. and G. Muhammad, *Cloud-assisted Industrial Internet of Things (IIoT) – Enabled framework for health monitoring.* Computer Networks, 2016. **101**: p. 192-202.
4. Palavalli, A., D. Karri, and S. Pasupuleti. *Semantic Internet of Things*. in *IEEE Tenth International Conference on Semantic Computing (ICSC)*. 2016.
5. Li, H., et al., *Enabling Semantics in an M2M/IoT Service Delivery Platform.* Proceedings - 2016 IEEE 10th International Conference on Semantic Computing (ICSC), 2016: p. 206-213.
6. Daniele, L., F. den Hartog, and J. Roes. *Created in Close Interaction with the Industry: The Smart Appliances REFerence (SAREF) Ontology*. in *Formal Ontologies Meet Industry: 7th International Workshop (FOMI)*. 2015.
7. EC, *Rolling plan for ICT standardisation 2017*. 2017.
8. Moreira, J.L.R., et al. *Towards IoT platforms' integration: Semantic Translations between W3C SSN and ETSI SAREF*. in *Semantics. Workshop SIS-IoT*. 2017.
9. Steinberg, J.S., et al., *2017 ISHNE-HRS expert consensus statement on ambulatory ECG and external cardiac monitoring/telemetry.* Heart Rhythm, 2017. **14**(7): p. e55-e96.
10. Fang, Y., C. Li, and L. Sun, *Design of an Early Warning System for Patients with Cardiovascular Diseases Under Mobile Environment.* Procedia Computer Science, 2016. **96**: p. 819-825.
11. Moreira, J., et al. *Improving the semantic interoperability of IoT Early Warning Systems: the Port of Valencia use case*. in *Enterprise Interoperability IX: I-ESA proceedings*. 2018.
12. Fortino, G. and R. Gravina, *Fall-MobileGuard: a smart real-time fall detection system*, in *Proceedings of the 10th EAI International Conference on Body Area Networks*. 2015, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering): Sydney, New South Wales, Australia. p. 44-50.
13. Gonçalves, B., V. Zamborlini, and G. Guizzardi, *An Ontological Analysis of the Electrocardiogram.* Journal of Communication, Information and Innovation in Health, 2009.
14. Gonçalves, B., G. Guizzardi, and J.G. Pereira Filho, *Using an ECG reference ontology for semantic interoperability of ECG data.* Journal of Biomedical Informatics, 2011. p. 126-136.
15. Cosío-León, M.A., et al., *The use of standards in embedded devices to achieve end to end semantic interoperability on health systems.* Computer Standards & Interfaces, 2018. **57**: p. 68-73.
16. HL7, *HL7 Version 3 Implementation Guide: Annotated ECG (aECG) R1, Release 2*. 2005.
17. Bond, R.R., et al., *A review of ECG storage formats.* International Journal of Medical Informatics, 2011. **80**(10): p. 681-697.
18. Solbrig, H.R., et al., *Modeling and validating HL7 FHIR profiles using semantic web Shape Expressions (ShEx).* Journal of Biomedical Informatics, 2017. **67**: p. 90-100.
19. Fortino, G., et al., *BodyCloud: A SaaS approach for community Body Sensor Networks.* Future Generation Computer Systems, 2014. **35**: p. 62-79.
20. Moreira, J., L.F. Pires, and M.V. Sinderen. *Semantic interoperability for the IoT: Analysis of JSON for Linked Data*. in *Enterprise Interoperability IX: I-ESA proceedings. BD4EI.* 2018.
21. Gyrard, A., M. Serrano, and G.A. Atemezing. *Semantic web methodologies, best practices and ontology engineering applied to Internet of Things*. in *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*. 2015.
22. Moreira, J.L.R., et al., *Towards ontology-driven situation-aware disaster management.* Journal of applied ontology, 2015. **10**(3-4): p. 339-353.
23. Moreira, J.L.R., et al. *Ontology-driven Conceptual Modeling for Early Warning Systems: Redesigning the Situation Modeling Language*. in *MODELSWARD*. 2017.
24. Moreira, J., et al. *Menthor editor: An ontology-driven conceptual modeling platform*. in *FOIS*. 2016. CEUR Workshop Proceedings.
25. Guizzardi, G., *Ontological foundations for structural conceptual models*. 2005: Enschede. p. 416.
26. Bera, P.a.P., Geert, *The effects of construct redundancy on readers' understanding of conceptual models.* JOURNAL OF DATABASE MANAGEMENT, 2017.