An Ontological Approach to Integrate Health Resources from Different Categories of Services

Cong Peng, Prashant Goswami, Guohua Bai

Department of Creative Technologies
Blekinge Institute of Technology
Karlskrona, Sweden

Email: cong.peng@bth.se, prashant.goswami@bth.se, guohua.bai@bth.se

Abstract—Effective and convenient self-management of health requires collaborative utilization of health data from different services provided by healthcare providers, consumer-facing products and even open data on the Web. Although health data interoperability standards include Fast Healthcare Interoperability Resources (FHIR) have been developed and promoted, it is impossible for all the different categories of services to adopt in the near future. The objective of this study aims to apply Semantic Web technologies to integrate the health data from heterogeneously built services. We present an Ontology Language (OWL)-based ontology that models together health data from FHIR standard implemented services, normal Web services and Linked Data. It works on the resource integration layer of the presented layered integration architecture. An example use case that demonstrates how this method integrates the health data into a linked semantic health resource graph with the proposed ontology is presented.

Keywords—Health data integration; ontology; FHIR; Semantic Web; Web service; eHealth; REST.

I. INTRODUCTION

Chronic diseases have become one of the main threats to people’s health [1]. The caring of chronic diseases requires long-term and periodical management by both the patients and healthcare staff. The high cost and inconvenience of the chronic diseases caring make it better for the patients to perform self-management [2].

The development of information and communication technologies makes it much more feasible for health self-management. Electronic Health Record (EHR) systems have been adopted by many healthcare providers. Portable medical devices are used by patients for self-monitoring physiological parameters [2][3]. Moreover, many people today use wearable devices and health applications to record and manage their health [4][5].

The aforementioned systems, devices and applications record a huge amount of health data about patients. The collaborative utilization of the various health data has the potential to support chronic disease patients in having more effective and convenient self-management [6][7]. Unfortunately, all these health data became data silos, which can only be utilized in their own places with very limited outside collaboration. One reason behind this situation is that the systems holding these health data are heterogeneously built.

Web services technologies have promoted the interoperability of various software applications running on distributed and diversified systems. However, the lack of common standards adoption still makes it problematic for integrating health data with heterogeneous data models.

Various works have been done on health informatics standards, among which the Fast Healthcare Interoperability Resources (FHIR) created by the Health Level Seven International (HL7) organization is regarded as the next generation of health information interoperability framework that combines the previous standards’ features [8][9]. It leverages common Web standards, which includes applying Representational State Transfer (REST) architectural style and JSON serialization format, besides the previous supported XML, as interfaces for health information exchange [10].

The FHIR working group is also working on the FHIR Linked Data module. It utilizes Semantic Web technologies include Resource Description Framework (RDF) and Web Ontology Language (OWL) to enhance its semantic expression capability and to facilitate inference and data linkage across datasets [8][10].

The work in this paper follows the path of FHIR to apply Semantic Web technologies and the REST resource model to integrate health data from different services as linked resources. We aim to link health services that adopted HL7 FHIR, implemented RESTful Web APIs or published as Linked Data, i.e., services with different levels of interoperability. The proposed approach is built on top of a semantic data aggregation method in [11]. The heterogeneous health data are modeled as conceptual information resources by using the Linked Health Resources (LHR) ontology. It makes the entire method a framework that aggregates health data from different sources and integrates them as health resources with semantics for upper level utilization.

The reminder of this paper is organized as follow. Section II introduces the previous works on integrating health data. Section III presents the method, which includes the health resource modeling, LHR ontology and the integration architecture. An example use case will be presented in Section IV to demonstrate the health data integration process by the proposed method. Finally, the paper will end with the conclusion section.

II. RELATED WORK

Many works have been done on integrating health data to support healthcare monitoring and decision making for either healthcare professionals or patients. In order to enable healthcare providers to remotely monitor and interpret health trends of diabetes patients, a method to integrate blood glucose data recorded from a patient-facing device to an EHR system was
presented in [2]. The integration was achieved by transmitting the glucose data to the device vendor’s iOS mobile phone application. It shares the data with the Apple HealthKit, which then sends the data to the EHR system. This solution is locked to the iOS platform since it depends on the Apple HealthKit.

V. Gay and P. Leijddekkers demonstrated a mobile application approach to aggregate health and fitness data to enable interoperability [5]. It was achieved by an Android application they developed with third-party partners to connect with wearable devices, EHR systems and other applications. A patient-centric mobile healthcare system that integrates data from body sensors was presented in [12]. The integration was implemented by leveraging a RESTful Web service on the application layer of the system to enable data sharing among applications.

There are also works done with Semantic Web technologies to enable health data integration with semantics. SENHANCE is a framework proposed by I. Pagkalos and L. Petrou to integrate patient self-reported health data on social network with hardware sensor observation data supported by Semantic Web technologies [3]. It models the former type of data as human sensor observations together with hardware sensor observations described by the ontology they proposed. B. Tilahun et al. presented a Linked Data-based system to retrieve and visualize heterogeneous health data in a flexible and reusable way [13]. The system utilized a set of Semantic Web technologies including RDF, Fuseki and SPARQL (SPARQL Protocol and RDF Query Language) for data representation, storage and query.

To integrate functionalities of different devices for supporting home-based care, an integration platform architecture was presented by Y. Trinugroho, F. Reichert and R. Fensli [14]. A smart home ontology that covered the modeling of person, device and context was proposed to enhance the reasoning process. To give semantics for the exchanged data between devices and systems in the home context, another smart home ontology was proposed in [15]. In addition, for the purpose of increasing the usability, the integration of smart home data with other data sources was also explored by applying Linked Data principles.

Despite there have been many works on using Semantic Web technologies to integrate health data, the integration between the ordinary health services and the services implemented with the next generation standard FHIR remains unexplored.

III. METHODS

The LHR framework intends to integrate 3 categories of health data sources: FHIR-enabled services, health services implemented Web APIs (satisfy certain REST constraints) and health related Linked Data. The 3 categories cover most available health related services and data, either for the current stage or the near future development trends.

A. The Nature of Health Resource

Today most of the health services that provide data access are providing Web APIs, which usually serve data in JSON or XML serialization format via HTTP methods. Most of these Web APIs claim to be RESTful services. Though in many cases they only follow a few of the REST constraints, which makes them actually not RESTful services. However, most of the services follow one of the fundamental elements of REST to organize the accessible data as resources, which is the unit of information in REST architectural style [16]. It makes them capable of being modeled together under the LHR framework.

In order to improve the implementability and to be more developer-friendly, the HL7 FHIR standard leverages the common Web technologies and concepts. FHIR therefore follows the REST architectural style as well, and is built upon a set of resources. Resource in the case of FHIR means “a collection of information models that defines the data elements, constraints and relationships for the objects relevant to healthcare” [8]. The objects that are modeled as resources include Patient, Observation, OperationOutcome and so on. Each resource is defined in a certain structure with references to other resources, and represented in XML, JSON and an additional RDF serializable format Turtle.

Linked Data is a practice of publishing structured and interlinked data with semantic meanings on the Web to make it a Web of data [17]. It was proposed by Tim Berners-Lee as an application of Semantic Web technologies. The Linked Data rules align well with some of the REST constraints. There have been many works on linking REST Web services and Linked Data [18][19]. Different levels of practices exist on publishing Linked Data according to the 5 star rating system developed by Berners-Lee [17]. A proper practice of publishing Linked Data should identify interlinked things (data items or real world entities represented on the Web) with HTTP Uniform Resource Identifiers (URIs) and serve corresponding information in RDF serializable formats or SPARQL query service.

Accordingly, we can regard anything identified by an HTTP URI as a resource, i.e., a node in an RDF graph. And the resource identified by the root path of an HTTP URI can be regarded as a service, i.e., a root node of an RDF graph.

![Figure 1. A simple overview of the integration process (a circle stands for a resource), resources are aggregated from services into semantic resource graphs, which are then integrated as a Linked Health Resource graph](image_url)
resource graph \( G(P_k) \). A health resource graph means a set of health information being modeled as resources with relationships, which maps to the RDF graph data model. Then \( G(P_k) = \sum S_i = \sum \sum R_{ij} \) denotes that a person’s health status can be described by sets of health resources from a collection of services. The sets of health resources can be represented in an integrated health resource graph. Figure 1 shows a simple overview of the process that resources from different services being integrated into a health resource graph.

Furthermore, a resource itself contains several objects:

- A resource may have interlinked (subordinate or related) resources \( R_{\text{link}} \).
- A resource may have data items \( D_i \) that describe certain status about a person, for example profile data, observation data, etc.
- A resource can be annotated by a set of \( n \) semantic tags \( T_r \), where \( n \in [0, \infty] \).

A resource can therefore be represented as a tuple \( R_j = (D_r, R_{\text{link}}, T_r) \).

C. Linked Health Resource Ontology

When we look into the very nature of the relationships among person, service and resource, we can see that a service as a unit also has resources that are interlinked, which is the same as a resource. It makes a service to be a resource. The same applies to a person, it then makes a person to be a resource as well. Therefore, we model the three types of objects all as sub-classes of class \( lhr:Resource \) under the LHR ontology, where the basic relationships among them being \( lhr:hasResource \) and \( lhr:isResourceOf \). Figure 2 shows the main classes and properties in the LHR ontology. It is built via the W3C recommendation standards, the RDF Schema (RDFS) data modeling vocabulary [20] and the OWL ontology language [21].

As we have stated, the class \( lhr:Resource \) is the super-class of all the involved resources. The class \( lhr:Service \) is a sub-class of \( lhr:Resource \), and represents a set of resources in a relatively separated territory to indicate a source of health data. The class \( lhr:Person \) represents a person who owns all the health resources. The class \( lhr:HealthResource \) represents the person instance’s all health resources that are interlinked from normal health Web services and FHIR implemented services. Usually Linked Data are not about only a single person, so resources from Linked Data is classified as the more generic class \( lhr:LDResource \).

To link the FHIR health-related resources, which are categorised as clinical resources in FHIR, we map those clinical resources into the LHR ontology. For instance, the \( fhir:Observation \) is mapped to \( lhr:FHIRObservation \), which is a sub-class of \( lhr:HealthResource \). Health resources from other normal Web services or Linked Data sets can also map from a class that have been defined in other ontologies or vocabularies to a sub-class of \( lhr:HealthResource \). Alternatively, the origin resource can instantiate directly from \( lhr:HealthResource \) if there is no embedded class.

Object properties around the class \( lhr:Resource \) and its sub-classes are the transitive property \( lhr:hasResource \) and its sub-properties include \( lhr:hasHealthResource \), \( lhr:hasLDResource \) and \( lhr:hasInterLinkedResource \). Among which, the property \( lhr:hasHealthResource \) has \( lhr:Resource \) as its domain and \( lhr:HealthResource \) as its range, the property \( lhr:hasLDResource \) has \( lhr:Resource \) as its domain and \( lhr:LDResource \) as its range.

The class \( lhr:DataItem \) is used to represent the data model of a \( lhr:Resource \) if it has no predefined model, and class \( lhr:ObservationData \) is used to represent the health data model of a \( lhr:HealthResource \). Resources from FHIR service have predefined data models in RDF, therefore, the object of which are mapped to an instance of \( lhr:ObservationData \). The relation is represented by a sub-property of the object property \( lhr:hasObservationData \).

With the sub-class relations and transitive sub-property relations defined in the LHR ontology, we can easily infer an instance of \( lhr:Person \) has a set of instances of \( lhr:HealthResource \) being linked together without complex inference rules.
D. Layered Architecture towards Integration

For integrating different health data into one ontological model with LHR, we need to firstly aggregate the health data together from different services. Figure 3 illustrates the layered architecture for integrating health data from different categories of services, as aforementioned that the LHR framework intends to integrate, into a LHR ontological model.

From bottom to top, the first layer is the data API layer. This layer simply requests data from services via APIs, e.g., Web APIs or Linked Data APIs. The retrieved data are then aggregated together into a semantic resource graph on the data aggregation layer.

In order to be effectively aggregated into the semantic resource graph, it is necessary for the normal Web services to have some simple semantic annotations with commonly used vocabularies. Besides the Semantic RESource Tagging (SemREST) in [22], JSON-LD context embedding to ordinary JSON representation [23] or other Semantic Web service annotation methods such as the ones presented in [24]–[26] could also be used to semantically annotate the resource.

As FHIR implemented services follow the FHIR standards, therefore, the data resources from these services come with standard structure and semantics. Some of the services even implemented FHIR/RDF representations. Data retrieved from Linked Data services are already RDF serializable if they were implemented properly. So the two categories of services can be aggregated into a semantic resource graph without much effort.

The semantic resource graph aggregated with the retrieved health data is then sent to the information resource integration layer. The LHR ontology is used to extract health resources from the resource graph for integration. One thing that needs to be noted is that the concept resource in the semantic resource graph from the data aggregation layer is slightly different from the lhr:Resource in LHR ontology. The former resource maps to a resource in a REST Web service, it contains functional meta information of its Web API. Only its representation unit will be integrated into the integration layer as an instance of lhr:HealthResource.

On top of the information resource integration layer there could be semantic APIs that utilize the health resource instances of lhr:Resource and the contained lhr:DataItem. Providing APIs on modelized resources in this manner has the potential to provide a unified interface to utilize the health data from different sources in the wrapped model for applications like health data visualization, analytics and so on. Alternatively, applications could also utilize health data in the form of semantic resource graph above the data aggregation layer. Thanks to all the resources are represented in RDF on both the aggregation layer and integration layer, it is therefore able to serve SPARQL queries directly.

IV. A USE CASE OF INTEGRATION

To demonstrate the LHR health data integration process, an example use case will be presented in this section. It is depicted as an idealized scenario where it leaves aside some details like the authentication and invocation of services. The demonstration prototype was implemented as two parts: the aggregation layer was implemented in Python with RDFlib package, and the integration layer was implemented in Java with Jena framework. The example use case that will be presented focuses on the integration layer since it is the main content of this paper. Therefore, the working process of aggregation is omitted here.

In a scenario that a patient named Alice has taken some medical examination at a healthcare center, which provides health data access via an HL7 FHIR implemented service. At the same time, she also has been using a Fitbit health band to track her daily activity and life style data for some time. In order to have a periodical comprehensive view of her health, she wants to integrate her blood glucose examination data from the healthcare center and her data of daily steps performed together.

For demonstration purpose, the data sample of blood glucose uses the FHIR official observation example [27]. Listing 1 shows the simplified resource of Alice’s blood glucose observation data in RDF/Turtle format. This resource is passed through the data aggregation layer to the integration layer directly since the FHIR observation resource is an instance of the class fhir:Observation that maps to lhr:FHIRObservation.

Next, we are going to integrate the daily steps performed data. Fitbit provides Web APIs to access the data recorded by their wearable devices. For integrating the data from Fitbit Web service into the LHR model, we choose to apply the SemREST method [11][22] to annotate its service description and aggregate the resource representation on the data aggregation layer. The generated semantic resource graph of Fitbit daily steps performed is partly shown in Listing 2 in RDF/Turtle format. It is then passed to the integration layer.
Listing 1. Simplified FHIR observation of blood glucose example represented in RDF/Turtle format. The observation resource contains the value, the effective date time and the LOINC code indicating the medical terminology.

```
@prefix fhir: <http://hl7.org/fhir/> .
@prefix loinc: <http://loinc.org/v2/1.5/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://hl7.org/fhir/Observation/f001> a fhir:Observation; 
  fhir:hasValue 9496 .
  fhir:hasDate 2018-01-14T09:30:10+01:00 .
  fhir:hasSystem /user/activities/steps/date/ .
  fhir:hasResource /user/activities/steps/ .
  fhir:hasComponent lhr:hasObservationData .
  fhir:hasDateTime 2018-01-12T01:00:00+01:00 .
  fhir:hasValueQuantity 5.5 .
  fhir:hasValueUnits mmol/l .

Listing 2. Simplified resource graph of Fitbit steps resource generated by SemREST and represented in RDF/Turtle format. The resource contains each day’s steps count, date time and the steps term in schema.org as a tag indicating the semantic meaning.

```
@prefix fitbit: <http://fitbit.com/namespace#> .
@prefix schema: <http://schema.org/> .
@prefix semrest: <http://semrest.org/vocab#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://api.fitbit.com/> a semrest:Representation ;
  semrest:hasResource fitbit:resource1 .
  fitbit:resource1 a semrest:Resource ;
  semrest:hasResponse fitbit:responsel ;
  semrest:hasMethod semrest:get ;
  semrest:hasTag schema:steps ;
  schema:activities ;
  semrest:urlTemplate */user/activities/steps/date/} 
  \[base-date]/\[end-date]/json* 

  fitbit:responsel a semrest:Response ;
  semrest:hasRepresentation fitbit:representation1 .
  fitbit:representation1 a semrest:Representation ;
  semrest:dataProvider [ semrest:hasValue "7430" ] ;
  semrest:hasTag schema:steps ;
  schema:dateTime "2018-01-12T" xsd:date ] ;
  semrest:hasValue "7352" ;
  schema:dateTime "2018-01-13T" xsd:date ] ;
  semrest:hasValue "4739" ;
  semrest:hasTag schema:steps ;
  schema:dateTime "2018-01-14T" xsd:date ] ;
  semrest:hasValue "9496" ;
  semrest:hasTag schema:steps ;
```

Listing 3. Example SPARQL query to retrieve integrated LHR health resources with their data value, date time and value type.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT * 
WHERE { 
  <@Alice> lhr:hasResource ?resource .
  ?resource lhr:hasDataItem ?dataItem .
  ?dataItem lhr:hasDataValue ?value . 
  ?dataItem lhr:hasDateTime ?dataTime . 
  ?dataItem lhr:hasValueType ?valueType .
}
```

In the integration layer, the coming resources are integrated using the LHR ontology. As mentioned before, the FHIR blood glucose observation resource is an instance of lhr:FHIRObservation, which is an equivalent class to fhir:Observation. For the ease of accessing data value of FHIR observation resources with LHR, the property fhir:Observation.valueQuantity is mapped to the corresponding object property lhr:FHIRObservation.valueQuantity, which is a sub-class of lhr:hasObservationData. And the fhir:Observation.effectiveDateTime is mapped to the corresponding data property lhr:hasObservationDateTime. Other objects could be accessed directly through the fhir:Observation instance, including the fhir:Observation.code, fhir:Observation.interpretation and fhir:Observation.referenceRange. Furthermore, for a more convenient utilization for upper level APIs, the object of fhir:Observation.dataValue could be mapped to lhr:DataItem’s value type to indicate the semantic meaning of the data value.

For the Fitbit daily steps performed resource, the semrest:Representation instance fitbit:representation1 in the resource graph from the aggregation layer is instantiated as a lhr:HealthResource. Since the steps data has been annotated as the property semrest:DataItem’s objects, so it can be mapped to the objects of lhr:hasObservation, a sub-property of lhr:hasDataItem, to this LHR health resource. The data value, date time and tag could also be mapped to the objects in the ontology for the upper level APIs.

The RDF of the integrated health resources is depicted graphically in Figure 4 with relatively important objects. Each square block represents an instance with its class and super-class. Instances are linked by object properties and data properties.

As one of the methods that can utilize the integrated health resources, a simple SPARQL query in Listing 3 is executed to query for Alice’s integrate health resources. The query result is presented in Figure 5. We can see that both the FHIR observation resource and the Fitbit resource are retrieved, since both the two types of resources were modeled as lhr:HealthResource directly or through sub-class relation. As the value, date time and value type have been mapped into the LHR ontology, so we can query them with the properties of lhr:DataItem. The value type of each data item is actually an URI. It not only works as the value type’s dereferencable URI for indicating the semantic meaning, but also integrates the resource under LHR with other Linked Data, where the
same URI were referenced.

In order to demonstrate the health data integration in a simple view, the example use case was performed far from a realistic way of managing health. However, it is capable to work in real cases since the data integration process is the same.

V. CONCLUSION

In this paper, we presented a method to integrate HL7 FHIR interoperability standard implemented health services and normal Web based health services with a few semantic annotations. The proposed OWL based LHR ontology modeled different categories of health resources together in a clear simple way for upper level unified utilization. Thanks to the application of Semantic Web technologies and the well aligned conceptual model with Linked Data, this method can also integrate health resources from the Linked Open Data on the Web without much effort.

With the capability to integrate different categories of health services, which usually contain health data from healthcare providers, consumer-facing products and health research, the presented method has the potential to support self-management of health especially for people with chronic diseases. To achieve a satisfied interoperable integration, it is ideal that the involved health services could implement interoperability standards such as HL7 FHIR. However, obstacles include services with legacy system problems make it impossible for all the services to realize in the near future. Diversity may exist among the different health services. The approaches like the one we proposed can act as a bridge to connect the heterogeneously implemented services. Integrating health data as an ontological model with semantic meaning is beneficial for upper level utilization in a more unified way.

Limitations, however, exist in this work. It still requires extra semantic annotation by the service providers for integrating the normal Web health services. Otherwise, it is unable to identify which part of the resource should be integrated. The more semantic annotations embedded in a resource’s representation, the more machine-understandable integration it will be, which can promote more useful applications. Manual ontology mapping is also needed for the resources from the data aggregation layer to be integrated on the integration layer.

For future works, we plan to evaluate this method with
more real world services for improvement and validation. In addition, the upper level semantic utilization APIs and applications will be explored for the collaborative utilization of health data.

REFERENCES


Copyright (c) IARIA, 2018. ISBN: 978-1-61208-675-0