Combining Personal Health Records and Relevant External Data Sources: A Way for Achieving New Outcomes for Personal Healthcare

Juha Puustjärvi
University of Helsinki
Helsinki, Finland
email: juha.puustjarvi@cs.helsinki.fi

Leena Puustjärvi
The Pharmacy of Kaivopuisto
Helsinki, Finland
email: leena.puustjarvi@kolumbus.fi

Abstract—**Personal Health Records (PHRs) have the potential to dramatically contribute to healthcare as they enable patient to become more involved and engaged in their care. However, PHRs are rather limited in that they assume all its content to be restricted on health-oriented personal data. Yet there are a lot of related data that are stored in other systems, and which use together with PHRs' data would produce outcomes that could not be achieved by functioning independently. Using these data sources together with PHRs' data we can achieve new outcomes. How it can be carried out by using modern Semantic Web technologies, such as Resource Description Framework (RDF), Web Ontology Language (OWL) and SPARQL, and our designated ontologies is the topic of this paper. We also introduce the notion of SPARQL-affinity domain, which allows the sharing of PHRs and other relevant data in a controlled way over the Internet.**

Keywords - Personal Health Records; Semantic Web; Open Data Sources; SPARQL

I. INTRODUCTION

Issues with combining heterogeneous data sources, under a single query interface have existed from the early 1980s, when computer scientists began designing systems for interoperability of heterogeneous databases. Nowadays, at the advent of the Semantic Web the issues with combining data sources is still equally relevant as the Semantic Web paradigm involves a broad set of modern technologies such as Resource Description Framework (RDF) [1], Web Ontology Language (OWL) [2] and SPARQL [3] that tackle these issues.

A relevant issue is how these technologies can be used in combining relevant external data sources with Personal Health Records (PHRs) [4]. A PHR is a record of a consumer that includes data gathered from different sources such as from health care providers, pharmacies, insures, the consumer, and third parties [5]. It includes information about medications, allergies, vaccinations, illnesses, laboratory and other test results, and surgeries and other procedures [6]. It is accessible to the patient and to those authorized by the patient.

A problem of current PHRs is that they assume all its content to be restricted on health-oriented personal data. However, there are a lot of related data, which use together with PHR data would produce outcomes that could not be achieved by functioning independently.

Examples of such PHR-related personal data sources include gyms, smart homes and personal note books. Gyms store data that is gathered by sensor and training equipment. Smart homes store a lot of data related to heating, air conditioning, and personal well fare such as weight measurements. Personal note books may include a variety of useful information concerning working hours, meals and location data.

Using these data sources together with PHRs' data we can achieve new outcomes: For example, a person may be interested to know his or her blood pressures when his or her weight had maximal and minimum values. Also, a person might be interested to know his or her cholesterol values grouped by his or her daily training hours.

There are also a lot of public data sources, which use together with personal data would produce outcomes that would not be achieved by using only personal data. For example, personal data may indicate the vaccinations of a person, while public data source can augment this information by more informal descriptions of the vaccinations. Further, as public data sources are increasingly linked among themselves according to the notion of the Linked Data [7], the data sources that can be reached from PHRs is increasing all the time. For example, the open data sources dealing with medicines or clinical guidelines [8] are particularly useful when used with PHR-data.

Especially, clinical guidelines will have a key role in self-care. They are documents with the aim of guiding decisions and criteria regarding diagnosis, management, and treatment in specific areas of healthcare [9]. They are based on an examination of current evidence within the paradigm of *evidence-based medicine*, which is one of the most important developments in the clinical use of information over the last decades [10]. Thus, the ability to reach clinical guidelines from PHRs enable patient to become more involved and engaged in their self-care.

However, a problem in combining external data sources with PHRs is data heterogeneity: there is a variety data models on which these data sources can be based on. For example, a data source may be a relational database or an XML-file. Further, the schemas of PHRs' XML-files may be ad hoc (e.g., the CCR standard of the American Society for Testing and Materials (ASTM) [11]) or based on the Reference Information Model (RIM) (e.g., the CCD standard of the HL7 [12]).

Our solution for this heterogeneity problem is the use of smart data. *Smart data* refers to data that is application-

independent, and part of a larger information ecosystem. Furthermore RDF [1] is the key for representing smart data.

RDF is a directed, labeled graph data format for representing information in the Web. It is not a data format, but a data model with a choice of syntaxes for storing data files [13]. In RDF, we can express facts with tree-part statements known as triples. The subject identifies the thing being described, predicate is a property name, and object is property value. So, each triple is like a little sentence that states a fact [14].

However, RDF in itself does not bring smartness. It depends on the expression power of the used vocabulary. By a vocabulary we refer to a set of ontologies, which formally specifies the used terms and their semantics. The key point here is that shared ontologies provide the ability of two or more systems to exchange information and to use the information that has been exchanged [15].

We will present ontologies that enable PHR system to interoperate with other relevant data sources. We illustrate these ontologies in a graphical way as well as in OWL. Data sources (RDF-files) are queried by SPARQL and these queries are processed by SPARQL processors [16].

The rest of the paper is organized as follows: First, in Section II, we introduce the notion of the SPARQL-affinity domain, which is the key concept in our designed system. The ontologies of the SPARQL-affinity domain are introduced in Section III. Then, in Section IV we present our used method for designing ontologies from XML-schemas. Especially we present how we have developed the Personal Health Ontology from the XML-schema of the Continuity of Care (CCD) documents. We present the Personal Health Ontology in a graphical way as well as in OWL. How we can query PHR-data and other external RDF-formatted data in one SPARQL query is considered in Section V. An example of presenting an ontology instance in RDF is given in Section VI. Finally, Section VII concludes the paper.

II. SPARQL-AFFINITY DOMAIN

SPARQL is an RDF query language to retrieve and manipulate data stored in RDF format. The name SPARQL is a recursive acronym for SPARQL Protocol and RDF Query Language, which is described by a set of specifications from the W3C [3]. SPARQL Protocol refers to the rules for how a client program and a SPARQL processor exchange SPARQL queries and results. There is a variety of SPARQL processors available for running queries against data both locally and remotely [16].

In our architecture, RDF formatted data stores that agree to work together for data sharing are called a *SPARQL-affinity domains* (Figure 1). Its data stores agree on a common set of policies such as how the data stores are accessed by web services, how users are identified, and how the access is controlled. However, the used policy is data store specific. For example, in the case of personal data, (such as with PHR-data, welfare data, and personal note book) a strict access policy is followed while in the case of public data sources (such as with public medicine data) no access control is needed. The access policy of smart home data can be defined case-by-case.
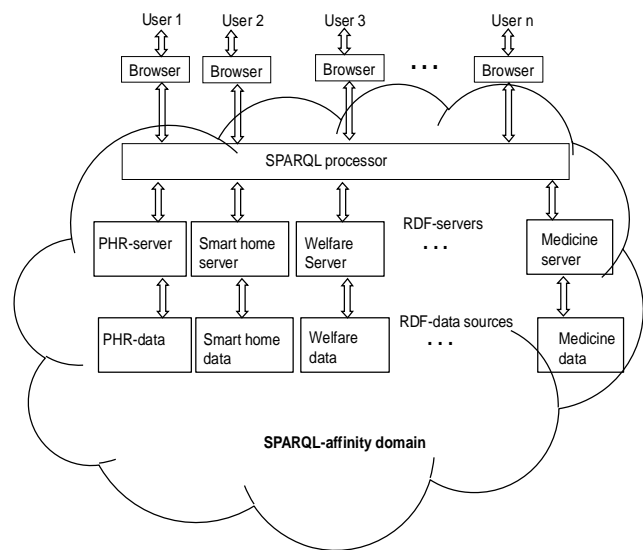

Figure 1. The architecture of the SPARQL affinity domain.

Each server provides a SPARQL endpoint [3]. A *SPARQL endpoint* is a web service that accepts SPARQL queries, runs the queries, and then returns the results. The way how multiple endpoints can be remotely processed in a SPARQL query is more detailed considered in Section V.

III. ONTOLOGIES IN SPARQL-AFFINITY DOMAIN

Each data source in the SPARQL-affinity domain is comprised of RDF-triples. These data sources are based on an ontology, i.e., these ontologies provide a vocabulary for these triples.

In computer science, an ontology is a general vocabulary of a certain domain, and it can be defined as "an explicit specification of a conceptualization" 17]. Essentially the used ontology must be shared and consensual terminology as it is used for information sharing and exchange.

Essentially ontology tries to capture the meaning of a particular subject domain that corresponds to what a human being knows about that domain [18]. It tries to characterize that meaning in terms of concepts and their relationships. It is typically represented as classes, properties, attributes and values. Depending on the generality level of conceptualization, different types of ontologies are needed [19]. Each type of ontology has a specific role in information sharing and exchange.

As an example consider a simplified Welfare Ontology, which is graphically presented in Figure 2. It comprises a vocabulary that a person can use in describing his or her personal welfare information. Hence, we do not assume that a person uses all the terms of the vocabulary (ontology). For example, datatype properties Father and Mother are included in the vocabulary, but the person does not have to give values for these properties. Neither the person needs class Swimming, if swimming is not included in his or her hobbies.
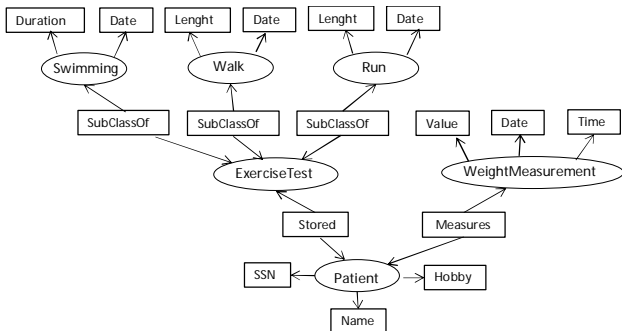
Figure 2.   A graphical presentation of a portion of the Welfare Ontology.

As an example of an ontology of a public data source consider the Medicine Ontology in Figure 3. This ontology can be used in storing information about the medicines and their manufacturers. For example, it provides links to data sources (ProductInfoUrl), i.e., to the web pages, that gives detailed information about medicines.
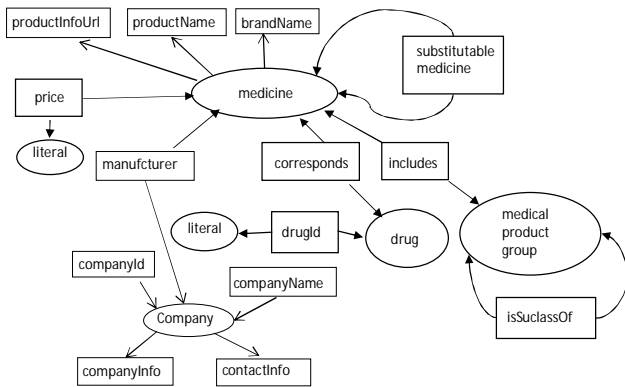

Figure 3.   A graphical presentation of a portion of the Medicine Ontology.

IV.   TRANSFORMING THE CCD-SCHEMA INTO PHR-ONTOLOGY

If an original data source of a SARQL-affinity domain is not in RDF, then we have to developed an appropriate ontology, and then transform the original data in the form, which is consistent with the developed ontology. For example, the data of most PHRs are based on the Continuity of Care Record (CCR) -standard [11] or CCD-standard [12], and therefore, we have developed an appropriate ontology, called the PHR-Ontology for these standards (XML-schemas).

Both CCR and CCD standards represent two different XML schemas designed to store patient clinical summaries. Both schemas are identical in their scope in the sense that they contain the same data elements such as demographics, medications laboratory results [20]. However, the structures the two XML schemas are quite different. Anyway the use of XML assures that the data contained in CCR or CCD documents can be expressed in multiple media formats that are friendly to both consumers and providers.

The CCD specification is a constraint on the HL7 CDA standard. The CCD standard has been endorsed by HIMMS (Healthcare Information and Management Systems Society Though) [21] and HITSP (Healthcare Information Technology Standards Panel) [22] as the recommend standard for exchange of electronic exchange of components of health information.

Although the original purpose of the CCD documents was to deliver clinical summaries between healthcare organizations, nowadays it increasingly used for other types of messages: it is increasingly considered as set of templates because all its parts are optional, and it is practical to mix and match the sections that are needed [23].

In transforming the XML schema of the CCD file to OWL-ontology we have used the following rules:

1. The complex elements of the XML-schema are transformed into OWL classes.
2. The simple elements of the XML-schema are transformed into OWL data properties such that the complex element is the domain of the da-ta properties.
3. The attribute of the XML-schema are transformed into OWL data properties.
4. The relationships between complex elements must be named and transformed to OWL object properties.

To illustrate this transformation consider the following example of a CCD document.

```
<SimplifiedCCDfile>
  <DocumentID>DOC_123</DocumentID>
  <Patient>
    <PatientID>AB-12345></PatientID>
    <PatientName>Tim Jones></PatientName>
  </Patient>
  <Medications>
    <Medication>
      <MedicationID>Medication.567</MedicationID>
      <DateTime>
        <ExactDateTime>2012-03-01TO12:00</ExactDateTime>
      </DateTime>
      <Source>
        <Actor>
          <ActorID>Pharmacy of Kaivopuisto</ActorID>
          <ActorRole>Pharmacy</ActorRole>
        </Actor>
      </Source>
      <Description>
        <Text>One tablet three times a day</Text>
      </Description>
      <Product>
        <ProductName>Voltaren</ProductName>
        <BrandName>Diclofenac</BrandName>
      </Product>
      <Strenght>
        <Value>50</Value>
        <Unit>milligram</Unit>
      </Strenght>
      <Quantity>
        <Value>30</Value>
        <Unit>Tabs</Unit>
      </Quantity>
    </Medication>
  </Medications>
</SimplifiedCCDfile>
```

Figure 4.   A simplified example of a CCD document.

In order to illustrate the transformation rules, let us consider the graphical OWL-ontology in Figure 5, which is derived from the elements presented in the document presented in Figure 4. In the figure, ellipses represent classes, and rectangles represent data type properties and object properties. Data type properties relate objects to datatype values while object properties relate objects to other objects.
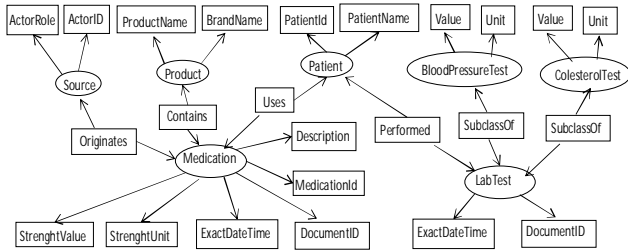

Figure 5.   A simple graphical PHR-Ontology.

The graphical ontology of Figure 5 is presented in OWL in Figure 6. Due to the space limits, we have omitted the specifications of the data properties such as PatientName, ProductName and BrandName.

```
<rdf:RDF
        xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-nsl#
        xmlns:rdfs=http://www.w3.org/2000/01/rdf-schema#
        xmlns:owl=http://www.w3.org/2002/07/owl#>
        <owl:Ontology rdf:about="ProfileCCDontology"/>
        <owl:Class rdf:ID="Patient/">
        <owl:Class rdf:ID="Medication/">
        <owl:Class rdf:ID="Source/">
        <owl:Class rdf:ID="Product/">
        <owl:Class rdf:ID="LabTest/">
        <owl:ObjectProperty rdf:ID="Uses">
                <rdfs:domain rdf:resource="#Patient"/>
                <rdfs:range rdf:resource="#aMedication"/>
        </owl:ObjectProperty>
        <owl:ObjectProperty rdf:ID="Contains">
                <rdfs:domain rdf:resource="#Medication"/>
                <rdfs:range rdf:resource="#Product"/>
        </owl:ObjectProperty>
        <owl:ObjectProperty rdf:ID="Originates">
                <rdfs:domain rdf:resource="#Medication"/>
                <rdfs:range rdf:resource="#Source"/>
        </owl:ObjectProperty>
</rdf:RDF>
```

Figure 6.   A simple PHR -Ontology in OWL.

Note that Figure 5 and Figure 6 represent only the portion of the PHR-Ontology that correspond the Medications section of the CDD. The whole PHR-Ontology is comprised of the integration of the ontologies derived from all sections of the CDD. In such integration we do not have to take care of semantic heterogeneity (i.e., one term is used in different meanings, or two terms are used in same meaning) as the all the elements in CDD documents are based on the HL7 RIM [24].

Note that querying PHR-data together with Medication data instead of querying only PHR-data we can achieve more expressive queries on patient's medication. For example, we can query the links of the web pages that provide information about the medications that are included in patient's medication. The ways multiple data sources can be queried in a SPARQL-affinity domain is the topic of the next section.

V.   QUERYING DATA SOURCES BY SPARQL

A typical SPARQL query specifies the pieces of information that meets the stated conditions. The conditions are described with triple patterns, which are similar to RDF triples but may include variables to add flexibility in how they match against the data.

SPARQL provides two ways for querying remotely: using FROM keyword or using SERVICE keyword [3]. In the former way the FROM keyword names a dataset to query that may be local or remote file. In the latter way, instead of pointing at an RDF file somewhere, a SPARQL endpoint is pointed. A SPARQL endpoint is a web service that accepts SPARQL queries, runs the queries, and then returns the result.

Federated Queries in SPARQL allow searching multiple datasets with one query [16]. For each dataset it is created a subquery which access datasets by using SERVICE keywords. That is, federated SPARQL queries make use of subqueries and SERVICE keywords. To illustrate this consider the federated SPARQL query presented in Figure 7, which accesses two data sets The query is based on the PHR-Ontology presented in Figure 5, and on the Medicine Ontology presents in Figure 3. Prefix *med* in the query refers to the PHR-Ontology while prefix *phr* refers to the Medicine Ontology.

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX phr: http://www.cs.helsinki.fi/PHROntology#>
PREFIX med: <http://www.cs.helsinki.fi/medicineOntology#>

SELECT ?drugId ?price
WHERE
{
  SERVICE <http://phrRegistry/sparql>
  { SELECT ?medicineId
   WHERE
   {
    Nancy Smith phr: uses ?medicineId
   }
  }
  SERVICE <http://medicineRegistry/sparql>
  { SELECT ?drugId ?price
   WHERE
   {
     ?medicineId med: corresponds med: drugId
     ?med: drugId substitutable_drug ?drug
     ?drug med: price ?price
    }
  }
}
```

Figure 7.   A simple federated SPARQL query.

The use case behind the query is the following: Nancy Smith is interested to know whether her medicines are substitutable with cheaper ones. Processing this query requires first to retrieve Nancy's used medicines from her PHR, and then querying the Medicine data set (a public RDF-formatted data source based on the Medicine Ontology).

As illustrated in Figure 7, the first subquery returns the medication identifications (medicineIds) of Nancy's medication, which in turn is the input parameter for the second subquery. This subquery first finds the active substance of Nancy's medicines, and then checks which other medicines include the same active substance (i.e., are substitutable). Finally, the main query outputs the medication identifications and the prices of these medicines.

This kind of cross-referencing feature is very useful in the SPARQL affinity domain as there is a variety of needs to cross-reference data from multiple data sources.

VI. REPRESENTING SMART DATA BY RDF

In order that RDF data can be represented and transmitted it needs a concrete syntax, which is given in XML, i.e., RDF statements are usually coded in XML. Hence, RDF inherits the benefits associated with XML. However, other syntactic representations (e.g., Turtle [25]) are also possible, meaning that XML-based syntax is not a necessary component of the RDF model.

One RDF description may contain one or more RDF statements about an object. For example, in Figure 8, the description concerning Mary Taylor's weight measurement (identified by "weightmeasurement100820151028") contains five RDF statements: the first states that its type in the Welfare Ontology is WeightMeasurement, and the second states that it measures Mary Taylor.

```
<rdf:RDF
    xmlns : rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns : po=http://www.helsinki.fi/Welfare_Ontology#>
    <rdf:Description rdf:about="weightmeasurement100820151028">
        <rdf:type rdf:resource="&po;WeightMeasurement"/>
        <po : Measures>Mary Taylor</po : Measures>
        <po : Date >10:08:2015</po:Uses>
        <po : Time >10:28</po:Time>
        <po : Value >68.7</po:Value>
    </rdf : Description>
</rdf:RDF>
```

Figure 8.   An instance of the Welfare Ontology in RDF.

VII. CONCLUSION

Internet has changed the way people work, bank and shop, but a similar change in health care has been small-scale. However, the use of Internet-based e-health tools is rapidly increasing. These tools cover many fields including electronic health records, personal health records, telemedicine, evidence based medicine, information therapy and disease management.

Still a problem is that the each e-health tool has its own interfaces and data sources. By integrating the e-health tools we can achieve two gains: simplify user interaction and

provide new more advanced services. In particular there are a lot of related data that are stored in other systems, and which use together with PHRs' data would produce outcomes that could not be achieved by functioning independently. Using these data sources together with PHRs' data we can achieve new outcomes. Further, the Semantic Web paradigm involves a broad set of modern technologies such as RDF, OWL and SPARQL that can tackle the issues with combining heterogeneous data sources.

In this paper, we have restricted on considering the SPARQL affinity domain as a technical infrastructure. However, to succeed the SPARQL affinity domain should not be considered just as a technical infrastructure but rather as ecosystems having many interconnected parts. Other key components of the whole ecosystem include governance regulations, financing and stakeholders. In our future work, we will restrict ourselves on analyzing the dependencies of these components.

REFERENCES

[1] "RDF – Resource Description Language". Available at: http://www.w3.org/RDF/ Retrieved: 2, 2016.

[2] "OWL – WEB OntologyLanguage", Available at: http://www.w3.org/TR/owl-features/ Retrieved: 2, 2016.

[3] "SPARQL Query Language for RDF", Available at: http://www.w3.org/TR/rdf-sparql-query/ Retrieved: 2, 2016.

[4] C. Angst, R. Agarwal, J. Downing, "An empirical examination of the importance of defining the PHR for research and for practice," Proceedings of the 41st Annual Hawaii International Conference on System Sciences. 2008.

[5] J. Puustjärvi and L. Puustjärvi, "Capturing Semantics for Personal Health Records" In Poc. Of the 6th International Conference on Information Technology and Applications (ICITA 2009), 2009, pp. 26-31.

[6] J. Puustjärvi, and L. Puustjärvi. "The Role of Care Pathways in Personal Health Records," In the proc. of the International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2015). 2015.

[7] DuCharme. Learning SPARQL. O'Reilly Media.

[8] N. Mulyar, M. van der Aalst, M. Peleg, "A pattern-based analysis of clinical computer interpretable guideline modeling languages," Journal of American Medical Information Association, 2007, vol. 14, no. 6, pp.781–787.

[9] M. Peleg, S. Tu, P. Ciccarese, J. Fox, "Comparing computer-interpretable guideline models," A case-study approach. Journal of American Medical Information Association, 2003, pp.10–58.

[10] T. Leaon, K. Kaiser, S. Miksch, "Free and open source enabling technologies for patient-centric,guideline-based clinical decision support," A survey. Methods of Information in Medicine, 2007, pp. 74–86

[11] "Continuity of Care Record (CCR) Standard", Available at: http://www.ccrstandard.com/ Retrieved: 2, 2016.

[12] CCD, "What Is the HL7 Continuity of Care Document?", Available at: http://www.neotool.com/blog/2007/02/15/what-is-hl7-continuity-of-care-document/ Retrieved: 2, 2016.

[13] J., Davies, F. Fensel, D. Harmelen. "Towards the semantic web: ontology driven knowledge management," West Sussex: John Wiley & Sons.2002.

[14] G. Antoniou, and F. Harmelen, "A semantic web primer," The MIT Press. 2004.

[15] J. Puustjärvi, and L. Puustjärvi, "Developing a Data Infrastructure for Patient-Centered Telemedicine",

International Journal on Advances in Life Sciences (IJALS). Vol. 6, No.3&4, 2014, pp. 229-239.

[16] "SPARQL Query Language for RDF", Available at: http://www.w3.org/TR/rdf-sparql-query/ Retrieved: 2, 2016.

[17] T.R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing." Padua workshop on Formal Ontology, March 1993.

[18] M. Daconta, L. Obrst, K. Smith, "The semantic web. :A Guide to the Future of XML, Web Services, and Knowledge Management," John Wiley & Sons. 2003

[19] J. Puustjärvi, and L. Puustjärvi. "Ontology-Based Competence Management in Pharmacy." International Journal of Adaptive, Resilient and Autonomic Systems (IJARAS). 4(2), 2013, pp 94-112.

[20] T. Benson, "Principles of Health Interoperability HL7 and SNOMED," Springer. 2010.

[21] HIMMS (Healthcare Information and Management Systems Society, http://himms.org/

[22] Healthcare Information Technology Standards Panel, http://hitsp.org/ Retrieved: 2, 2016.

[23] K. Boone, " The CDA Book,". Springer. 2001.

[24] R. Dolin, L. Alschuler, C. Beerb, P. Biron, S. Boyer, E. Essin, T. Kimber,. "The HL7 Clinical Document Architecture", J. Am Med Inform Assoc, 2001:8(6), pp. 552-569.

[25] "Turtle - Terse RDF Triple Language," Available at: http://www.w3.org/TeamSubmission/turtle/. Retrieved: 2, 2016.