2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)

Semantic Lifting and Reasoning on the Personalised Activity Big Data Repository for Healthcare Research

Hong Qing Yu[1], Xia Zhao[2], Zhikun Deng[1], Feng Dong[1]

[1]Department of Computer Science, University of Bedfordshire, Luton, UK
[2]School of Computing and Digital Technology, Birmingham City University, Birmingham, UK

Abstract— The fast growing markets of smart health monitoring devices and mobile applications provide opportunities for common citizens to have capability for understanding and managing their health situations. However, there are many challenges for data engineering and knowledge discovery research to enable efficiently dealing with the data that are collected from heterogonous devices and applications with big volumes and velocity. This paper presents a research work in MyHealthAvatar project, which developed a comprehensive semantic driven knowledge discovery framework based on the integrated data from multiple data resources. The framework applies cloud-based hybrid database architecture of NoSQL and RDF repositories with introductions of semantic oriented data mining and knowledge lifting algorithms. The major aim of the research is to enhance the knowledge management and discovery capabilities and efficiency to support further accurate health risk analysis and lifestyle summarization.

Index Terms — Knowledge management, Semantic Web, Data Engineering, Data integration, Healthcare.

1 Introduction

THIS paper illustrates an innovative framework that manages and integrates multiple health-related data resources from wearable sensors, mobile and web applications for discovering people's health knowledge, which will exert influence on the future direction of people's self-care empowerment, disease prevention and importantly promote better lifestyles. With fast development of health-oriented smart sensor devices, mobile applications and social network technologies, various different kinds of information related to human health can be more efficiently collected. In many cases, these data give us big advantages to shift medical care from institutions to the home environment, and to transform healthcare from a system that is largely reactive – responding mainly when a person is unwell – to one that is much more proactive in supporting patients in self-management [1]. On the one hand, recent research evidence shows that patients with chronic conditions who are more actively involved in their own healthcare receive better health outcomes [2]. On the other hand, understanding a person's current condition and tracking their behaviours over time provides many opportunities to enable improving their overall health and well-being. In addition, self-management skills can be developed and strengthened, even among those who are initially less confident, less motivated or have low levels of health literacy [3]. One of the key factors of success in self-healthcare empowerment is to allow the patient to gain valuable and understandable knowledge from their own data, bringing them tangible benefits.

There has been indeed growing interest in the 'initiative of self-monitoring', evidenced by the sharp market expansion in life-logging devices and apps. These sensors are capable of constantly monitoring personal health behaviours and activities (e.g. walking, calories, heart rate, and diet), leading to unprecedented opportunities in self-care. Correspondingly, significant research effort has started to harvest and integrate the sensors for long-term health data collections – examples include MyHealthAvatar [4] and MyLifeHub [5]. Such a long-term data collection is extremely valuable to individualised disease prediction and prevention, and to promoting healthy lifestyles. Although many data can be collected from different devices or applications, it is currently still quite difficult for people to discover correlations about themselves. Even with tools developed for advanced devices like the Withings scale and Fitbit pedometer to track their daily weight and step count, it is still not possible to get clear pictures of conditions and trends between the two or to see how they interact on specific days of the week, weekends versus weekdays, month to month, etc. without exporting the data into understandable knowledge [6, 7]. To achieve knowledge understanding and management, the highly heterogeneous and dynamic nature of the data brings new challenges. We note here three challenges that we would like to address in this paper.

(1) Efficiently storing, querying and integrating the big dataset that collected from heterogeneous data resources.

(2) Providing scale and easily understandable knowledge discovery mechanism to observe significant factors from a big and day-to-day data repository.

(3) Semantically representing the personalised health knowledge to support both human understanding

978-1-5386-3066-2/17 $31.00 © 2017 IEEE
DOI 10.1109/iThings-GreenCom-CPSCom-SmartData.2017.125

and further machine-based dynamic knowledge analysis and reasoning.

We understand that these are not all the challenges, but the proposed framework focuses on addressing these three highlighted challenges by providing a comprehensive infrastructure that can efficiently manage a large integrated dataset from heterogeneous data collections and semantically discover and represent the significant health related knowledge (event). The framework combines two types of modern data repositories – NoSQL database and Semantic RDF database with semantic knowledge reasoning engine between them, which is applied as foundation to our research contributions.

The rest of the paper is organized as follows. Section II discusses our research background, research motivations and the literature review on related work. Section III explains the overall architecture of the proposed framework. Section IV introduces the ontology defined for the semantic data retrieval and reasoning. Section V focuses on the illustration of the knowledge discovery algorithm. Section VI explains the semantic lifting and reasoning processes in detail. Section VII provides the case study and evaluation results of our research. Section VIII finally draws the conclusion and outlines potential future work.

2 Related Work and Motivation

2.1 Project background

Health-related data integration work has been focused by several research projects in the last few years. For example, the MyHealthAvatar (MHA) framework is a proof–of-concept EU-funded 3 million euros project for providing a digital representation of patient health status. It is designed as a lifetime companion for individual citizens that facilitates the collection of, and access to, long-term health-status information that includes citizens' social and sensor data, together with major data resources from traditional healthcare organisations.

2.2 Healthcare-related ontology

There are many medical ontologies that aim to support different medical research tasks of clinical research, trail investigation and biomedical investigation.

Clinical Trial Ontology (CTO) and Ontology of Clinical Research (OCRe) [17] are developed to describe methods for binding to external information standards (e.g. BRIDG) and clinical terminologies (e.g. SNOMED CT [19]). These standards allow the indexing of research studies across multiple clinical trials and observational studies, interventions/exposures, outcomes, and health conditions. With such indexing, investigators interested in the evidence pertaining to a particular question (e.g., what is the effect of A on B in people with C) will be able to locate relevant research studies more easily across disparate data sources.

The Ontology for Biomedical Investigations (OBI)1 is an open access, integrated ontology for the description of biological and clinical investigations. OBI provides a model for the design of an investigation, the protocols and instrumentation used, the materials used, the data generated and the type of analysis performed on it. In OBI the common formal language used is the Web Ontology Language (OWL).

TABLE 1
Data Integration Source and Type

Data Collection	Sources	Data type
Steps	Fitbit/Moves/ MHA app	Count
Travelling & activity type	Fitbit/Moves/ MHA app	Minutes & transport type
Location	Moves/ MHA app	Coordination
Diet	MHA app	Calories & food category
Health Profile	MHA app – profile input	PHR like records
Weight	Withings scale	Grams
Body fat	Withings scale	Grams
Blood pressure	Withings BPM	mmHg
Slept hours	Fitbit/Withings	minutes
Awoken times	Fitbit/Withings	Count
Social activity	MHA app – calendar	Description

PRotein Ontology (PRO)2 has been designed to describe the relationships of proteins and protein evolutionary classes (ontology for ProEvo), to delineate the multiple protein forms of a gene locus (ontology for protein forms), and to interconnect existing ontologies. PRO provides an ontological representation of protein-related entities by explicitly defining them and showing the relationships between them. Each PRO term represents a distinct class of entities (including specific modified forms, orthologous isoforms, and protein complexes) ranging from the taxon-neutral to the taxon-specific.

Disease-Treatment Ontology (DTO) [31] is developed to model and represent treatment information found in the abstracts of medical articles. The aim of the DTO is to develop an automatic extraction system to extract treatment information from medical abstracts retrieved from the Medline database, to support information retrieval, question answering, summarization, and knowledge discovery. The purpose of the ontology is to serve as a knowledge base to store the extracted information and support these functions.

Translational Medicine Ontology (TMO) [18] is developed as a unifying ontology to bridging the gap between different terminologies of chemical, genomic and proteomic data with disease, treatment, and electronic health records. TMO is to build semantic links between traditional patient health record (PHR) ontologies and the semantic knowledge based on linked data cloud such as DO [20] and SNOWED CT. However, the TMO still focuses on managing the knowledge of patients' formal health record data without considering the user's daily activity data, which is also important to discovery healthcare knowledge for the individual patient, or healthy user.

3 Overall Architecture

[1] http://obi-ontology.org/page/Main_Page

[2] http://pir.georgetown.edu/pro/pro.shtml

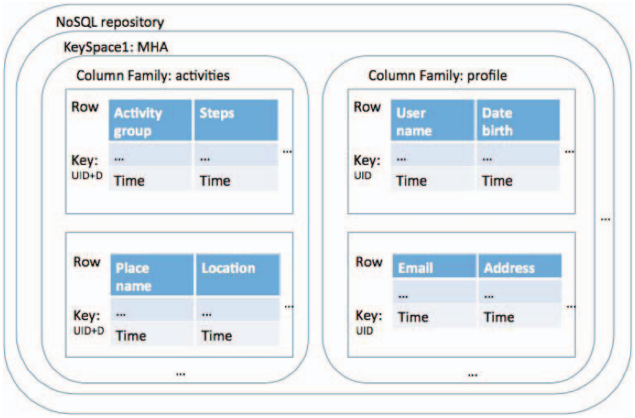

Fig. 1. NoSQL data model design examples

3.1 Data integration

Collecting and meaningfully integrating heterogeneous data resources is a longstanding problem in data management and engineering research areas. In our research, we collect desired data from multiple data resources including mobile applications (Moves), wearable sensors or digital measuring devices (Fitbit [21, 22] and Withings3) and MHA platform. Each different data resource provides different and useful data information, as Table 1 shows. The data collection process applies Web API technologies following the OAuth security protocol4. Whenever the user logs in to the MHA platform, the data from other devices can be synchronized into the system.

Fig. 1 shows the detail designed NoSQL database structure that bases on the column and key query data storing mechanism.

Each column family stores a group of rows that contains a set of individual columns in a specific data structuring requirement. For example, one row in the 'activities column' groups all the data columns that store activity type, step counting and duration data elements. The other row in the family can store the places information that the user has travelled to or plans to visit. The 'profile' column family completely focuses on managing basic user profile information such as name and contact.

3.2 Data mining process

Although the NoSQL database provides us a much faster data retrieval foundation, the data size of the integrated data pool is still large and is increasing every day. Moreover, the integrated data contains much noisy data in the data streams, especially the data collected from sensor devices. For example, the location coordination data are updated frequently according to just a couple of meters difference in the user's movement (even if the user is actually still inside the 'same' place). Therefore, the major purposes of a data mining process are (1) to provide a much smaller amount of data that is more significant and meaningful to understand the user's conditions, and (2) to allow efficient storage of data in the semantic layer for supporting further advanced knowledge discovery and reasoning. We define the 'event' concept to refer to a data group that describes a fact derived from the integrated data repository. The events require discoverability based on the available data resources, which covers two aspects:

Significant activity events include travelling to unusual or healthcare places, sport exercises, high calory consumption activity and social activity.

Physiological (symptom) events mainly refer to well defined symptoms such as low/high blood pressure, unusual heart rate, poor sleeping and significant weight/fat changes.The algorithms for mining the events are detailed in Section 5.

3.3 Semantic layer

The semantic layer comprises three components: Semantic repository, Semantic lifting engine and Semantic reasoning engine.

Virtuoso RDF repository has been deployed on our private cloud server to store the semantic triples and OWL ontology schema. Virtuoso automatically provides SPARQL endpoint and JDBC update connections to client applications and server-side developers.

Apache Jena RDF semantic reasoning framework is applied to develop our semantic reasoning engine. Three sets of SWRL based rules are defined in order to achieve our reasoning aims, which are described in Section VI.

4 MHA H-Event Ontology

The core concepts of the MHA H-event ontology includes 10 major terms, as Fig 2 presents. The ontology extends TMO terminologies with some existing semantic concepts from well-known domain ontologies and our defined personal activity and treatment terms. Some important terms are:

- **Event** is defined as the same as **TMO.Processual_entity** is a super concept to classify an interesting event that is related to the health of an individual user. The event is the super class of (discover a) **Symptom**, (taking a) **Treatment**, (diagnosed a) **TMO.Disease_progression** and (having a) **Activity**. Each event associates with a particular time point on the user's time.TemporalEntity. In addition, Event is the central point of the whole ontology, which can be detected from the data mining layer.
- **Disease_progression** reused from TMO is a subclass of the Event concept and presents the medical situations that were diagnosed in the past according the user's timeline or are a potential risk for the user. The Health Condition, Treatment and Symptom concepts structure a triangle relation that could be very valuable knowledge for an individual user or a group of users.
- **Risk** defined by **ICO** is used as a concept to evaluate the possibility or progress levels to a

[3] http://www.withings.com/

[4] http://oauth.net/2/

particular health condition.

- **Lifestyle** is imported using the Intention concept in the MWLA ontology[5] that defines 25 lifestyle instances. Since MWLA is still a live EC project (CARRE[6]), the numbers of the lifestyle definitions can be enriched in the future.
- **Knowledge** is an association term to the Event class that means that each item of knowledge is related to an instance of a discovered event such as an activity, a symptom, a treatment or a disease progress.

In this paper, we focus on the knowledge discovery on the sides of activity and symptom that can be dynamically mined from the monitoring data. However, the designed ontology is more general and comprehensive and can be used for integrating real clinical data such as PHR.

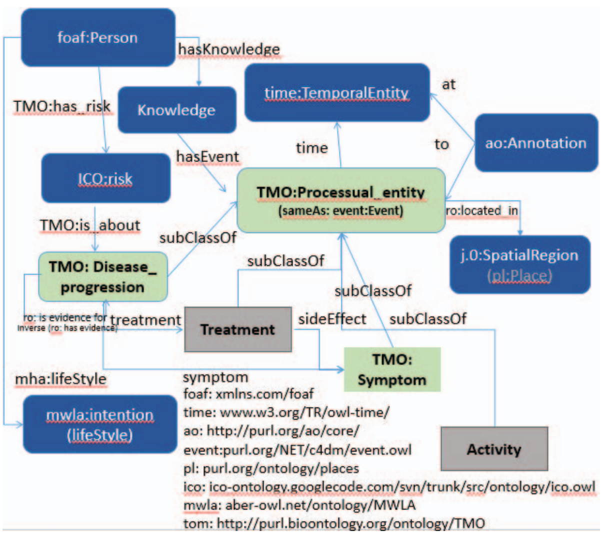

Fig.2. Top layer MHA H-event Ontology

5 Event Mining Algorithm

5.1 Significant activity event mining

SLAi is defined as the significant level of the *i*th activity detected from the lower-level data stream and Equation 1 is the calculation function:

$$SLAi = \prod Wi_j = Wi_{at} \cdot Wi_{loc} \cdot Wi_{dur} \qquad (1)$$

Where Wiat presents the activity type weight of the ith activity, Wiloc presents the activity location weight of the ith activity and Widur presents the activity duration weight of the ith activity. Therefore, the final level score is the ∏ of the three weight values. The range of each weight is [0.1, 1], therefore 0< SLAi ≤1.

Table 2 shows the weight value distributions of activity type and location type:

TABLE 2
WEIGHT VALUES FROM 0 TO 1 FRO DIFFERENT ACTIVITIES AND LOCATIONS

Activity type	Weight	Location type	Weight
home/work	0.1	home/ work place	0.2
walk	0.3	shop/ restaurant	0.4
transport	0.6	entertaining / sports/ social places	0.6
social	0.8	transport	0.8
exercise/ healthcare	1	other place	1/times been the place in the month +1

The activity duration weight is defined as Equation 2:

$$Wi_{dur} = TD \cdot (1/36000) \qquad (2)$$

Where TD is the duration (seconds) of the activity and 36,000 is the number of seconds in 10 hours.

Finally, the significant threshold for lifting the activity as semantic knowledge to the semantic repository is set as:

$$SLAi \geq 0.02$$

The weight definition is defined based on the analysis results of the real data and the desired scenarios. Table 3 shows the lowest boundery of some classical activity events.

5.2 Symptom (Physiological) event mining

The physiological event mining methods are defined based on medical measurement guidelines. At the moment, we concentrate on detecting four symptoms: high/low blood pressure, unusual heart rates. All these three symptoms are well-defined in medical guidelines.

For example, the blood pressures have systolic blood pressure that measures how hard the heart's left ventricle contracts to circulate blood through the body. Diastolic blood pressure measures the pressure in the blood vessels when the heart's chambers are relaxed and filling with blood. UK National Health Service (NHS) guidelines[7] indicates that normal adult blood pressure should be between 90/60 and 140/90, where the top (first) number isthe systolic pressure and the diastolic is the bottom (second) number. In addition, readings higher than 140/90 can be defined as high blood pressure and lower than 90/60 as low blood pressure.

The other mining methods are defined here based on similar UK NHS guidelines. Heart rate range should generally be in [60, 100], otherwise it is too slow or too fast. The sleep hours should generally be between six and nine hours.

[5] https://bioportal.bioontology.org/ontologies/MWLA
[6] https://www.carre-project.eu/
[7] http://www.nhs.uk/NHSEngland/thenhs/about/Pages/overview.aspx

6 SEMANTIC LIFTING AND REASONING

6.1 Semantic lifting

The semantic lifting process is a generalization of RDF triples based on the proposed ontology, which includes two steps of semantic mapping: domain mapping and property mapping with range assignment.

Step 1: Domain mapping

In the first step the domain matching algorithm is applied to identify the domain element; this is the simplest algorithm in these three steps. According to our JSON structure composing the summary data analysis, only activity type or symptom name elements are suitable candidates for the domain that can be lifted as subject elements of the instance RDF triples. If the element is under the activities JSON structure, then a URI will be generated and specified as an Activity class defined in the OWL ontology. For example, if the activity is classified as running, then the URI will be http://myhealthavatar.org/activity/running.

A similar process is generated for the symptom event.

Step 2: Property and range mapping

According to the JSON input structure, the property and range mapping are performed together based on the pre-defined mappings in Tables 4.1 and 4.2.

6.2 Semantic reasoning

The final goal to have the data lifted into the semantic repository is to enable mining the data further to discovery hidden knowledge about the user, getting the benefit of the smaller but more machine understandable data representations – RDF triples based on well-defined semantic ontology. we have developed a semantic reasoning engine using the Jena semantic framework that supports SPARQL and SWRL (Semantic Web Rule Language) programming and can be integrated into a Virtuoso RDF repository. The former is based on SPARQL queries that can define the reasoning formulas at the ontology level (T-box). The latter is based on SWRL rules that cannot be specified at the ontology level, rather at the instances level (A-box). To explain these two different reasoning processes, we represent two reasoning scenarios here to illustrate how reasoning can be applied for inference of lifestyle pattern and links to certain health conditions or symptoms.

Example 1: Travel/long commute lifestyle for the past month.

Definition: Long transport (more than 2 hours) activity events have been lifted in to the RDF at least four times in the last month. The reasoning process will be:

1) Construct (SPARQL query) the last month activity event RDF memory-model based on the ontology retrieved from the triple storage.
2) Specify the SWRL rule or SPARQL query for reasoning. In this example, Code 1 is used as the SPARQL query.

TABLE 4.1

JSON structure syntax	Mapped property defined in the ontology	Range value
ranking value	mha: rank	(0,1]
duration,	mha: time	Seconds
destination	mha: located in	Annotation text or unknown
distance	mha: distance	Metres
step	mha: step	Count
activity_group	mha: hasEvent	Activity type

```
PREFIX mha: <http://myhealthavatar.org/ontology/>
SELECT (COUNT(?numberOfLongTravelling) AS ?howMany ?p)
WHERE { ?p mha:hasEvent ?e .
      ?e rdf:type mha:transport .
      ?e mha:time ?d .
        FILTER (?d >= 7200)
}
HAVING ( ?howMany > 4 )
```

CODE 1

Code 1 specifies a query that will return transport times (COUNT ?howMany) a month if the transport activity events have been detected more than 4 times (use ASK query to filter ?howMany > 4) in a targeted month.

If the query returns a value, then it means the user satisfies the defined reasoning query. Then we can construct the semantic links between the person to the Travel term defined in the MWLA.

The steps of 2) and 3) can be integrated by combining Construct update query to the model as one step (Code 2).

```
PREFIX mha: <http://myhealthavatar.org/ontology/>
mwla: < http://purl.bioontology.org/ontology/MWLA>

CONSTRUCT (?p mha:lifestyle mwla:Travel)
SELECT (COUNT(?numberOflongTravelling) AS ?howMany ?p)
WHERE { ?p mha:hasEvent ?e .
      ?e rdf:type mha:transport .
      ?e mha:time ?d .
        FILTER (?d >= 7200)
}
  HAVING ( ?howMany > 4 )
```

CODE 2

The similar processes can be applied to detect other lifestyle, such as lack of activity or lack of exercise and so on.

Example 2: SWRL-based health-condition risk alarm reasoning.

Definition: If a person lacks activity and has age > 60, then there is a risk of high blood pressure.

The rule can be defined as Code 3 in the Jena rule engine.

```
[rule: (?p mha:lifestyle mwla:noActivity), (?p foaf:age ?i), greaterThan(?i, 60) -> (?p tom:has_risk ?x), (?x tom:is_about ?d), (?d rdf:type tom:High_Blood_Pressure)]
```

CODE 3

8 Conclusion

In this paper, we have illustrated and explored an advanced knowledge data management and discovery framework that works on the big data from the multiple data resources of smart sensors and applications in the personal health domain. The paper has demonstrated how different resources data has been integrated in our framework. In addition, a MHA H-event ontology has been defined to map the NoSQL model to become knowledge, which relies on the two domain-specific event-mining algorithms. These two algorithms have been developed to contribute to lift significant knowledge to the semantic layer in order to filter the most important information and scale down the size of the data for further data analysis. This process can be adopted to different domain applications for application to different datasets and specific knowledge discovery tasks. Our evaluation results show that the process very successfully addresses the issue of generating scaled semantic knowledge bases.

Finally, the paper introduced the semantic reasoning process that can further reason on the knowledge bases to provide valuable information for health condition analysis and summarize an individual's lifestyle features and relations to their health states. The reasoning processes can be defined in two ways: SPQRQL queries and SWRL rules. The explorations suggest the semantic reasoning is much more powerful for dealing with complex knowledge discovery scenarios than traditional SQL-based knowledge discovery processes in a big data context.

References

[1] Wagner, E.H., Chronic disease management: what will it take to improve care for chronic illness?, Effective Clinical Practice, vol 1, no 1, pp 2–4, 1998.

[2] Greene, J. and Hibbard, J.H., Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. Journal of General Internal Medicine, vol 27, no 5, pp 520, 2012.

[3] Hibbard, J.H. and Greene, J., What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs, Health Affairs (Millwood), vol 32, no 2, pp 207–14, 2013.

[4] H. Kondylakis, M. Spanakis, S. Sfakianakis, V. Sakkalis, M. Tsiknakis, K. Marias, Z. Xia, H. Q. Yu, F. Dong, "Digital Patient: Personalized and Translational Data Management through the MyHealthAvatar EU Project", International Conference of the IEEE Engineering in Medicine and Biology Society of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, Milan, Italy.

[5] P. Yang, M. Hanneghan, J. Qi, Z. Deng, F. Dong and D. Fan, Improving the Validity of Lifelogging Physical Activity Measures in an Internet of Things Environment, *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM),* Liverpool, 2015, pp. 2309-2314.

[6] Li, I., Dey,A. and Forlizzi, J. (2011) Understanding My Data Myself: Supporting Self-Reflection with Ubicomp Technologies. In Proc UbiComp 2011.

[7] Li, I. Personal Informatics and Context: Using Context to Reveal Factors that Affect Behavior. PhD Thesis. 2011.

[8] The Top Challenges in Big Data and Analytics, Lavastorm analytics, 2013, http://www.gsma.com/membership/wp-content/uploads/2013/07/The-Top-Challenges-of-Big-Data-Analytics.pdf

[9] Hu, B., Carvalho, N., Laera, L., and Matsutsuka. T., Towards big linked data: a large-scale, distributed semantic data storage. In Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services (IIWAS '12)

[10] Garg, M. K., Kim, D., Turaga, D. S. and Prabhakaran, B., Multimodal analysis of body sensor network data streams for real-time healthcare. In Proceedings of the international conference on Multimedia information retrieval (MIR '10). ACM, New York, NY, USA, 469-478.

[11] Bauschlicher, D., Bauschlicher, S., and ElAarag. H., Framework for the integration of body sensor networks and social networks to improve health awareness. In Proceedings of the 14th Communications and Networking Symposium (CNS '11). Society for Computer Simulation International, San Diego, CA, USA, 19-26.

[12] K. Tollmar, F. Bentley and C. Viedma, "Mobile Health Mashups: Making sense of multiple streams of wellbeing and contextual data for presentation on a mobile device," Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on, San Diego, CA, 2012, pp. 65-72.

[13] Diego Calvanese and De Giacomo, Giuseppe and Maurizio Lenzerini and Daniele Nardi and Riccardo Rosati, Data Integration in Data Warehousing, Int. J. of Cooperative Information Systems, vol. 10, no. 3, pp. 236-271, 2001.

[14] G. Wiederhold, Mediators in the architecture of future information systems, IEEE Comput. 25, 3 (1992) pp. 38-49.

[15] J. D. Ullman, Information integration using logical views, Proc. ICDT'97, LNCS (Springer-Verlag, 1997) pp. 19-40.

[16] Yang, H. and Li, W., An ontology-based approach for data integration in regionally interoperable healthcare systems. In: 11th International Conference on Informatics and Semiotics in Organisations (ICISO 2009), 11-12 Apr 2009, Beijing, China, pp. 93-96.

[17] Sim I, Tu SW, Carini S, Lehmann HP, Pollock BH, Peleg M, Wittkowski KM. The Ontology of Clinical Research (OCRe): an informatics foundation for the science of clinical research. J Biomed Inform. 2014 Dec;52:78–91.

[18] Luciano JS, Andersson B, Batchelor C, et al. The Translational Medicine Ontology and Knowledge Base: driving personalized medicine by bridging the gap between bench and bedside. Journal of Biomedical Semantics. 2011;2(Suppl 2):S1. doi:10.1186/2041-1480-2-S2-S1.

[19] Lee, Dennis et al. "Literature Review of SNOMED CT Use." Journal of the American Medical Informatics Association : JAMIA 21.e1 (2014): e11–e19. PMC. Web. 11 Apr. 2016.

[20] Schriml, Lynn Marie et al. "Disease Ontology: A Backbone for Disease Semantic Integration." Nucleic Acids Research 40.Database issue (2012): D940–D946. PMC. Web. 11 Apr. 2016.

[21] Fitbit Ltd, Fitbit Healthy Futures Report, 2013, http://www.trajectorypartnership.com/wp-content/uploads/2014/02/Fitbit-Healthy-Futures-Report-September-2013.pdf

[22] MACKINLAY, Molly Zellweger. Phases of Accuracy Diagnosis: (In) visibility of System Status in the Fitbit. Intersect: The Stanford Journal of Science, Technology and Society, [S.l.], v. 6, n. 2, jun. 2013. Date accessed: 27 Apr. 2016.

[23] DATASTAX Corporation, Introduction to Apache Cassandra – White Paper, July 2013. http://www.datastax.com/wp-content/uploads/2012/08/WP-IntrotoCassandra.pdf