Design and Development of a Linked Open Data-Based Web Portal for Sharing IoT Health and Fitness Datasets

Roberto Reda
University of Bologna
Department of Computer Science and Engineering
Bologna, Italy
roberto.reda@unibo.it

Antonella Carbonaro*
University of Bologna
Department of Computer Science and Engineering
Bologna, Italy
antonella.carbonaro@unibo.it

ABSTRACT

The huge amounts of self-tracked health data collected by Internet of Things (IoT) fitness devices offer important opportunities to the research community. If properly exploited, IoT health and fitness datasets can help to gain valuable insights into the human health in order to provide better healthcare.

However, IoT health data come from a variety of different heterogeneous sources and in proprietary formats, which means that they require an integration process, normally manually done by domain experts, in order to be analysed. This task is not only significantly time consuming but in many cases, error prone.

In this study, we designed and developed a web platform for collecting and publishing IoT health and fitness datasets according to Linked Data principles. We leveraged the IFO ontology and the Semantic Web technologies to make the IoT health and fitness datasets freely available to the community in a shared, semantically meaningful, easily discoverable, and reusable manner.

The system introduced in this article shows that Semantic Web technologies can be a viable and comprehensive solution for describing, integrating and sharing heterogeneous IoT datasets, thus overcoming the issues of data silos that nowadays dominate the IoT landscape.

CCS CONCEPTS

• Information systems → Information integration; **Resource Description Framework (RDF)**; **• Applied computing → Health informatics**; *Health care information systems*;

KEYWORDS

health informatics, semantic web, lod, iot, wearable devices

ACM Reference Format:
Roberto Reda and Antonella Carbonaro. 2018. Design and Development of a Linked Open Data-Based Web Portal for Sharing IoT Health and Fitness Datasets. In *International Conference on Smart Objects and Technologies for Social Good (Goodtechs '18), November 28–30, 2018, Bologna, Italy.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3284869.3284890

*Corresponding Author

Goodtechs '18, November 28–30, 2018, Bologna, Italy

ACM ISBN 978-1-4503-6581-9/18/11…$15.00
https://doi.org/10.1145/3284869.3284890

1 INTRODUCTION

In recent years, the advances in consumer-oriented Internet of Things (IoT) healthcare technologies have lead to the proliferation of wearable devices (including fitness trackers and smartwatches), mobile apps, and related services available on the market.

A wide range of individuals are now able to continuously capture and digitally store a variety of personal health data (PHD) such as physical activity, heart rate, blood pressure, body temperature, blood glucose, and sleep patterns.

Potentially, the IoT wellness devices could give users direct access to personal analytics that can contribute to improve their health, facilitate preventive care, and aid in the management of ongoing illness. Health and fitness data collected by smart devices can provide better insights of everyday behaviour and lifestyle, and can fill in gaps in more traditional clinical data collection systems [11]. Moreover, there seems to be an increasing willingness for individuals to share their PHD with others [33]; the Quantified Self (QS) movement is a notable example of this trend [37].

The sharing of the enormous amount of self-tracked health information, daily collected by users, presents an important opportunity for researchers to develop novel methods to deliver high quality care and ignite innovative projects which could have the potential to revolutionise the healthcare sector [11].

IoT wearable datasets have already been successfully used to study the relationship between sleep and mood problems [32], to detect unhealthy eating and exercise levels [29], and predict human behaviour [27].

However, most of the time IoT data come from a variety of different heterogeneous sources and are represented with their own proprietary format depending on the device manufacturer. This diversity and variety which characterises the IoT health and fitness datasets along with their huge volume make sharing and integration more difficult. Data heterogeneity is still an open issue challenge that needs to be addressed to fully exploit the potential of the IoT data.

Semantic Web (SW) technologies, based on common standards, offer opportunities to cope with the semantic data heterogeneity that hampers the integration and distribution of datasets drawn from diverse sources sharing the same context [28] [41][39].

For instance, ontologies, which constitute a formal conceptualisation of a specific domain, provide a common terminology that gives data a well-defined meaning that allow interconnection and reuse in ways other than as originally implemented or intended, independently of the type and the format of the data [15, 18, 19]. Moreover, SW technologies are not only useful for converting scattered health data into valuable aggregated information but also for sharing them [14, 16, 17].

Linked Data (LD) [12] is a set of best practises for publishing and exposing data as resources on the Web and interlinking them with semantically related datasets using SW technologies. LD is offering a worthwhile alternative to the isolated and heterogeneous data silos which dominate the IoT landscape since are based on standardised formats and interfaces.

Strictly related to the concept of LD, Linked Open Data (LOD) is a term which refers to the application of the LD principles to Open Data (i.e., data that can be freely used and distributed); a classic example of LOD collection is DBpedia [8].

In this paper, we present the design and development of a LOD portal which aims to become a reference point for collecting, publishing and sharing IoT health and fitness datasets in structured format, so that IoT data can be accessed and reused by domain experts, scientists and the web community with no restrictions by any form of licensing or patent.

Unlike other portals for sharing PHD, such as Open Humans [4] or Kaggle [3] which collect and redistribute users' raw data (i.e., data serialised in unstructured and semi-structured formats) only, a novel aspect of our work consists in providing a semantic representation of the IoT datasets. The ontology-based approach we adopted addresses the problem of data provided in heterogeneous formats by clarifying what the data describe, thus facilitating the integration, exploration and the analysis of the datasets and promoting innovative ways to use these data creatively.

The reminder of this paper is organised as follow. In Section 2 we present existing works for sharing PHD gathered from IoT devices and we compare them with our proposal. Section 3 reviews the core of SW technologies and LD. Section 4 details our work, the design and development of a web portal for sharing IoT health and fitness datasets according to LD principles. A brief discussion of the results achieved to illustrate the added value of having IoT data in a harmonised and machine readable format is provided in Section 5. Current limitations of the proposed system and future works conclude the article in Section 6.

2 RELATED WORKS

Within a context of scientific research and big data analysis, the IoT health data can only reach their full potential if they can be collected from multiple sources by many individuals, in huge volumes. Here, we introduce some of the existing platforms which provide public access to IoT health datasets.

Open Humans [4] is an open source project which aims to make more health-related data available for scientists and also help volunteers make that data more accessible to researchers. The online web portal allows users to upload, store and share their personal data such as genetic, social media data, activities and health data gathered through IoT devices.

Kaggle [3] is a large and diverse data community and provides a platform for predictive modelling and analytics competitions. Kaggle adopts a crowd-sourcing approach to collect large datasets from companies and users, including IoT wearable data.

PhysioBank [5] is a rich repository of physiological health-related signals datasets such as health-monitoring or vital signs, often associated to demographics information.

ResearchKit [7] is an open source framework introduced by Apple that allows medical researchers gather health data directly from users' smartphones to carry out clinical studies.

However, while these platforms are important initiatives to publish and share IoT health datasets online, data can only be retrieved in unstructured or semi-structure formats through proprietary Application Programming Interfaces (APIs), thus information integration, comparison, and reuse are very difficult. Moreover, little metadata is normally added to the datasets, and often vocabularies and data formats are inconsistent. This makes finding, assembling, and normalising these datasets time consuming and prone to errors.

A public access repository for sharing IoT health and fitness datasets expressed in a machine readable format (i.e., structured data), such the one we propose in this paper, can enable researchers and domain experts to find, share, and integrate information more easily. To our knowledge, a LOD portal specifically designed for sharing IoT health datasets has not yet been deployed.

Similar projects (not specific for the IoT data) have already successfully employed SW technologies for sharing datasets in life science. A notable example is Bio2RDF [10] which is a biological database that uses SW technologies to provide interlinked data gathered from the most important public bioinformatic databases. Another noteworthy example is BioPortal [38] which is a repository of over 300 biomedical ontologies. CardioSHARE [42] is framework for querying and performing analysis on distributed clinical data on heart diseases.

3 LINKED DATA AND SEMANTIC WEB TECHNOLOGIES

The term LD refers to a set of best practises for sharing and interlinking structured data and knowledge on the Internet by using standard web technologies [12]. The primary goal of the LD initiative is to make the Web not only useful for publishing documents, but also for sharing and interlinking single pieces of data. The movement is driven by the idea that the SW technologies facilitating the data sharing, integration, and analysis on a global scale could revolutionise the way we manage knowledge just like the Web revolutionised information sharing and communication over the last two decades.

Technologically, the core idea of LD is to use the Internationalised Resource Identifiers (IRIs) [36] to univocally identify arbitrary entities and concepts. Information about entities referred by IRIs can be simply retrieved by dereferencing the IRI over the HTTP protocol. Data about entities and concepts are then represented through the Resource Description Framework (RDF) [35] language. RDF is a standardised data model which uses graphs to represent information and facts by means of triples in the form subject, predicate, object. Whenever a web client resolves an IRI associated to a triple's subject of a resource, the corresponding web server provides an RDF description of the identified entity, these descriptions can contain links to other RDF graphs in the triple's object. Whenever an application resolves a predicate IRI, the corresponding server responds with an RDF Schema (RDFS) [20] or a Web Ontology Language (OWL) [23] definition of the link type, that is a vocabulary or an ontology. Ontologies are a key aspect of the SW since they enable interoperability among different systems by providing an agreed-upon terminology such as the basic terms and relations in a domain of interest, and as well as rules how to combine these terms. Because the Web of Data is based on standards for the identification, retrieval, and representation of information and knowledge, and scattered entities are interconnected by links, it is possible to crawl the entire data space, fuse data from different sources, and provide expressive query capabilities over aggregated data, similarly to how a local database is queried today. For this purpose, the Simple Protocol and RDF Query Language (SPARQL) [24] is the standard language for querying, combining and consuming structured data in a similar way SQL does this by accessing tables in relational databases.

Nowadays, the great majority of IoT health and fitness datasets are accessible only in human-readable formats such as HTML pages or property data formats, therefore users must have proprietary software to access the data. Since LD is exclusively based on open web standards, data consumers and domain experts can use generic tools to access, analyse, and visualise data. Moreover, LD make use of ontologies to formally define the meaning of entities and resource so that they do not limit the ability of machines to process data automatically.

4 METHODS

4.1 Overview

The aim of our project was to design and develop a LOD-based web portal in order to collect health and fitness data gathered from consumer health IoT devices, and make them freely available on the Web. For the design process of the system we mostly followed the detailed set of recommended practices for creating and publishing LD sources in the Health Care and Life Sciences (HCLS) domain as described in [30].

Our platform is capable of (1) collecting IoT fitness data manually entered by users or automatically retrieved from remote repositories, (2) integrating and storing IoT datasets semantically annotated according to a reference ontology, (3) visualising information through a customisable dashboard, and (4) sharing datasets adhering to LD principles. We developed our system adopting a four-tier architecture. A layered architecture makes the various parts of the system independent and logically separated, and single components, replaceable and upgradeable. For instance, at the moment, data can be entered to the system manually by users or downloaded automatically from the Cloud, the layered architecture would allow us to add a third method for collecting data (e.g., directly from devices) without affecting the rest of the system.

As can be seen in Figure 1, the four-tier architecture consists of the following layers: (1) the data retrieving layer, (2) the data processing layer, (3) the service layer, and (4) the presentation layer. The data retrieving layer collects IoT datasets from users or automatically from remote servers. The data processing layer transforms the IoT raw data in semi-structured formats into an RDF graph, datasets are thus semantically annotated according to a reference ontology and stored within a NoSQL database (i.e., a triplestore). The service layer controls data access and bridges the clients to the system via service protocols. The presentation layer allows users to interact with the system using either the web based dashboard or the SPARQL endpoint. We developed the web portal using JavaServer Pages (JSP) [2] as back-end technology.

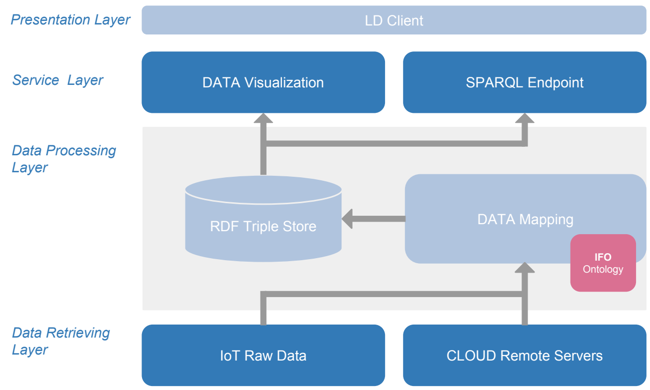

Figure 1: LOD system architecture with all the four layers from the data retrieving layer to the presentation layer.

More details about the various system components, with more focus on the role of SW technologies, are given throughout the next subsections.

4.2 Data Retrieving

Gathering and integrating in an homogenised way the huge volume of IoT health information is an extremely challenging task because data are spread across different platforms in heterogeneous formats (data silos). Datasets can be manually entered by users to our system or can be automatically retrieved from cloud storage systems since many vendors allow to obtain the data via their servers through public APIs.

Most IoT fitness vendors grant access to data stored on their servers using authentication mechanisms such as the Open Authorisation (OAuth) [25]. OAuth is an open standard and provides external applications secure delegated access to a server on behalf of the owner. OAuth specifies a process to authorise access to the resources without sharing the user credentials. Our system allows users to automatically retrieved data from Fitbit and Nokia Health servers. Once the user has given the permission to access health data on his behalf, the system can periodically download data without further user intervention.

IoT fitness servers do not rely on a standard for exchanging data. Without exception, each vendor defines its own specific proprietary API interfaces. Server-specific software has to be written to retrieve the data once the access authorisation has been granted.

The most common serialisation formats used for IoT health and fitness data are the Extensible Markup Language (XML), the Comma Separated Value (CSV) and the JavaScript Object Notation (JSON). The data formats promoted by the Health Level Seven (HL7) standardisation group seem to be ignored and not taken into account by any IoT fitness producer. Once the IoT data is retrieved manually of automatically from remote servers, a copy is maintained within an offline repository for archiving purposes.

4.3 Data Processing

In order to share the heterogeneous IoT datasets following the principles of LD, a unified representation for the different concepts as well as a formal semantics representation to these data is required.

4.3.1 Data Model. Ontologies provide a common understanding of the domain knowledge and increase data interoperability among different applications. Ontologies allow to formally describe the semantics of terms and items of the IoT fitness datasets and give explicit meaning to the provided information. Besides sharing a common understanding, another relevant benefit of a formal semantics is the support for machine-readability, which facilitates information integration and enables the application of intelligent approaches such as semantic search and automated reasoning.

Our system mostly relies on the IoT Fitness Ontology (IFO) [34] for the semantic annotation of datasets. The IFO ontology is a domain level OWL ontology specifically designed to formally represent the most common and important concepts within the domain of the IoT fitness devices and wellness appliances.

The IFO ontology is modelled around the concept of episode. An episode is a generic event that can be detected and registered by an IoT fitness device, concrete events such a blood pressure reading or a walking training session are specialisations of the generic episode and are organised in a hierarchical structure within the ontology.

To each episode is always associated a time stamp and a numeric value with the related unit of measurement. These data are essential because they allow to numerical quantify the object of the event and give it a temporal collocation and duration useful for reasoning processes. Moreover, several other metadata can be added to a single episode such as geographic coordinates (location is becoming a basic attribute for health data [9]) or user personal annotations.

The abstraction mechanism of episode allows us to describe low level concepts such as a single device sensor measurement (e.g., 36°C body temperature) as well as high level concepts like 20 minutes of hard run followed by 5 minutes of a moderate walk within the same ontology.

4.3.2 Data Transformation. LD recommendations require datasets to be published in RDF format. Since IoT datasets are retrieved in semi-structured formats (i.e., CSV, JSON, XML) a process to convert them to RDF is needed. In our system, during the conversion process, besides the ontology mapping, also well-formed IRIs are generated and assigned to entities within the datasets.

For the mapping process we chose to employ the RDF Mapping language (RML) [21] and the RML Processor for its execution.

RML is a declarative source-independent mapping language which allows to express customised mapping rules for converting heterogeneous resources into RDF graphs according to a reference ontology in an integrate and interoperable fashion. RML extends the W3C standard R2RML [6] which is defined to express customised mappings only from data stored in relational databases. RML keeps the mapping definitions as in R2RML but encompasses broader set of possible input sources.

This characteristic of RML makes it particularly useful within the IoT fitness context because, as already stated, different vendors use different data and serialisation formats to represent and store information about the same concept. For instance, Fitbit stores data about the body weight in JSON format, Nokia Health serialises the same type of data in CSV format while Apple Health does the same in XML. RML simplifies the development of a mapping specification for the same concepts since the definition of the triple structure has to be specified only once and can be reused across other sources in the same or different formats. RML mapping specifications can be generated in a semi-automatic way [22] or manually. Several mapping specifications for different devices are already available and ready to use within our system and users can also add their own specifications.

As soon as the triples are generated they are loaded to the triplestore.

4.4 Data Sharing

IoT health and fitness datasets transformed into a well-organised LD structure are exposed, on our portal, through a public SPARQL endpoint. A SPARQL endpoint essentially enables users or software application clients to query the RDF data via the SPARQL language.

From a SPARQL query perspective, once the data is represented as RDF and exposed via a SPARQL endpoint, the different storage modalities become irrelevant. However, SW systems, such as LD portals, usually build on triplestores as their main data storage. Triplestores provide data management and data access via APIs and query languages for RDF data.

For our system, we used Fuseki [1] as triplestore and SPARQL endpoint.

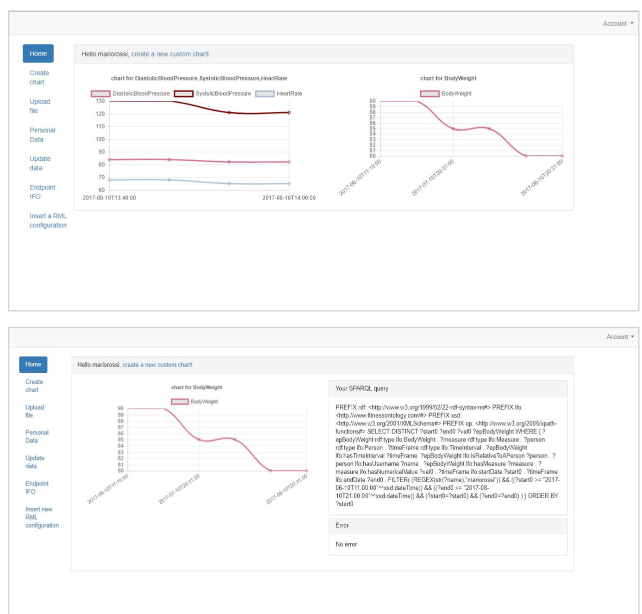

Figure 2: Two screen shots of the web-based dashboard showing some vital signs data charts and a customised body weight chart generated through a user defined SPARQL query.

4.5 Data Visualisation

Information visualisation is an important component of LOD portals since it increases accessibility of LD-based systems [26]. The main objective of information visualisation is transforming and presenting data into a visual representation, in such a way that users can explore and use the data.

Our system allows users access their personal data through a web-based visualisation dashboard which provides multiple views of their integrated datasets.

Health data representation methods must be flexible in order to cover the needs of users with different backgrounds and requirements. RDF data model offers unique opportunities since it enables to bind data to visualisations in unforeseen and dynamic ways [13]. For instance, when an information visualisation technique requires certain data structures to be present, we can derive and generate these data structures automatically from reused vocabularies or semantic representations, in this way we are able to realise a largely automatic visualisation workflow.

To exploit the flexibility offered by RDF we made the dashboard highly customisable by allowing expert users to define the information to be displayed on charts using a user-made SPARQL query (Figure 2). Federated queries are also possible within the dashboard.

Since writing a SPARQL query is a challenging task for nontechnical users, we provided the dashboard with several preset queries for visualising common information in the form of time series such as the heart rate or the blood pressure readings.

5 RESULTS

The main result of our work is a system which is able to integrate data from multiple heterogeneous IoT health and fitness sources and expose them in structured format so that they can be accessed and queried in a uniform way using standard language. Data visualisation in a personalised manner is also possible through a web dashboard.

To test our system we collected data about body weight, blood pressure and heart rate using three different IoT fitness devices.

The choice of the devices was based on the different serialisation formats they adopt to store the collected data.

Body weight data was generated and uploaded to the Nokia Health server through the dedicated smartphone app since Nokia Health give access to the datasets in CSV format. The blood pressure readings were partially uploaded to the Nokia Health server and partially stored on a smartphone within the Apple Health app, in the latter case data is stored in XML format. Heart rate information was collected through a Fitbit wristband (in JSON format) and an iOS smartphone app [31].

Data from Apple Health were manually entered to our system while body weight (Nokia Health) and heart rate readings (Fitbit) were automatically downloaded by our system from their respective remote servers. All the IoT datasets provided have been correctly uploaded and correctly transformed to structured data.

We used the web dashboard to visually explore the homogenised datasets and SPARQL for performing some statistical operations over the homogenised datasets.

It is noteworthy to underline that such queries have operated regardless of how the original data were stored and represented in the IoT sources (i.e., CSV, JSON and XML in our case). Moreover, we could have analysed our experimental datasets together with some other data from another LOD portal in the same way, since once the data is represented as RDF and exposed via a SPARQL endpoint, the different storage modalities become irrelevant from a SPARQL query perspective.

A similar task without the support of SW technologies would have been non trivial to formulate and execute, a federated queries "on the fly", as the one suggested, virtually impossible.

6 CONCLUSION

6.1 Discussion

The enormous amount of self-tracked health information collected by users through smart fitness devices, offers important opportunities to the research community.

The process of gathering and integrating data from scattered IoT sources is normally done manually by researchers and domain experts. This process is not only cumbersome but also significantly time consuming and in many cases, error prone.

An effective and efficient exploitation of the IoT health and fitness data requires methods for accessing, integrating, interpreting and analysing datasets from multiple distributed sources in a unified way.

We designed and developed an LD-based web platform which is capable of collecting and retrieving IoT health and fitness datasets from users and SQ enthusiasts in order to make them freely available to the research community. Our system can convert heterogeneous IoT raw data collected by a multitude of different devices into RDF graphs. The homogenised datasets are stored in structured format and exposed publicly via a SPARQL endpoint for accessing and querying.

However, we did not limit ourselves to merely convert the raw data into RDF and publish them online. In our system, we also leveraged the IFO ontology to make the IoT health and fitness datasets available in a shared, semantically meaningful, easily discoverable, and reusable manner.

Additionally, we provided the web portal with data visualisation capabilities so that the platform can also be used as a personal health record (PHR) system. Users can visualise and explore their integrated datasets through a customisable web dashboard.

To our knowledge, the LOD portal presented in this paper, is the first of its kind specifically designed to gather and share IoT health and fitness datasets.

In this study, we have also demonstrated that SW technologies can be a viable and comprehensive solution for describing and integrating the heterogeneous IoT health and fitness data. In particular we have showed that the LD initiative may offer unprecedented opportunities for exposing the information collected by IoT devices inasmuch LD rely on structured RDF graphs that can be queried uniformly via SPARQL. LD enable the fusion of local and public data in a very powerful way, thus overcoming the main issues of data silos.

6.2 Limitations

However, there are some limitations to this work. Firstly, the sample queries and the amount of data we used to test the system is relatively small, and a more robust and rigorous evaluation along several dimensions (e.g., performance, query results, robustness, devices supported) is required before it can be deployed. Secondly, sophisticated policies and practices that relate to privacy of health information have not been taken in consideration. Even if our system shares information only in anatomised form, when two or more sources of personal data are combined risk of revealing a person's identity increases significantly [40]. Finally, although the IFO ontology covers a vast set of IoT fitness concepts, describing the datasets more robustly with domain-specific, additional ontological vocabularies and interlinking with more ontologies is recommended.

6.3 Future Works

For future research, additionally to address the limitations aforementioned, we plan to implement an integrated search engine for supporting a better data discovery and access which can potentially make the LOD portal more interesting and usable by researchers and healthcare professionals.

Moreover, we intend to leverage the reasoning and inference mechanisms of the SW technologies and the IFO ontology to enhance the dashboard by providing it with more advanced data analysis capabilities specific for the IoT health and fitness datasets.

REFERENCES

[1] [n. d.]. Apache Jena Fuseki. https://jena.apache.org/documentation/fuseki2/. ([n. d.]). Last Accessed: 2018-08-16.

[2] [n. d.]. JavaServer Pages Technology. http://www.oracle.com/technetwork/java/index-jsp-138231.html. ([n. d.]). Last Accessed: 2018-08-16.

[3] [n. d.]. Kaggle. https://www.kaggle.com/. ([n. d.]). Last Accessed: 2018-08-16.

[4] [n. d.]. Open Humans Project. https://www.openhumans.org/. ([n. d.]). Last Accessed: 2018-08-16.

[5] [n. d.]. PhysioBank. https://www.physionet.org/physiobank/. ([n. d.]). Last Accessed: 2018-08-16.

[6] [n. d.]. R2RML: RDB to RDF Mapping Language. https://www.w3.org/TR/r2rml/. ([n. d.]). Last Accessed: 2018-08-31.

[7] [n. d.]. ResearchKit. https://developer.apple.com/researchkit/. ([n. d.]). Last Accessed: 2018-08-16.

[8] D Abián, F Guerra, J Martínez-Romanos, and Raquel Trillo-Lado. 2017. Wikidata and DBpedia: a comparative study. In *International KEYSTONE Conference on Semantic Keyword-Based Search on Structured Data Sources*. Springer, 142–154.

[9] Nancy Andes and Jack E Davis. 1995. Linking public health data using geographic information system techniques: Alaskan community characteristics and infant mortality. *Statistics in medicine* 14, 5-7 (1995), 481–490.

[10] François Belleau, Marc-Alexandre Nolin, Nicole Tourigny, Philippe Rigault, and Jean Morissette. 2008. Bio2RDF: towards a mashup to build bioinformatics knowledge systems. *Journal of biomedical informatics* 41, 5 (2008), 706–716.

[11] Matthew J Bietz, Cinnamon S Bloss, Scout Calvert, Job G Godino, Judith Gregory, Michael P Claffey, Jerry Sheehan, and Kevin Patrick. 2015. Opportunities and challenges in the use of personal health data for health research. *Journal of the American Medical Informatics Association* 23, e1 (2015), e42–e48.

[12] Christian Bizer, Tom Heath, and Tim Berners-Lee. 2011. Linked data: The story so far. In *Semantic services, interoperability and web applications: emerging concepts*. IGI Global, 205–227.

[13] Josep Maria Brunetti, Sören Auer, Roberto García, Jakub Klímek, and Martin Nečaskỳ. 2013. Formal linked data visualization model. In *Proceedings of International Conference on Information Integration and Web-based Applications & Services*. ACM, 309.

[14] Antonella Carbonaro. 2009. Collaborative and semantic information retrieval for technology-enhanced learning. In *Proceedings of the 3rd International Workshop on Social Information Retrieval for Technology-Enhanced Learning (SIRTEL 2009), Aachen, Germany.*

[15] Antonella Carbonaro. 2010. Improving web search and navigation using summarization process. In *World Summit on Knowledge Society*. Springer, 131–138.

[16] Antonella Carbonaro. 2010. Towards an automatic forum summarization to support tutoring. In *Technology Enhanced Learning. Quality of Teaching and Educational Reform*. Springer, 141–147.

[17] Antonella Carbonaro. 2010. WordNet-based summarization to enhance learning interaction tutoring. *Journal of e-Learning and Knowledge Society* 6, 2 (2010), 67–74.

[18] Antonella Carbonaro and Rodolfo Ferrini. 2007. Ontology-based video annotation in multimedia entertainment. In *Consumer Communications and Networking Conference, 2007. CCNC 2007. 4th IEEE*. IEEE, 1087–1091.

[19] Antonella Carbonaro and Rodolfo Ferrini. 2008. Personalized information retrieval in a semantic-based learning environment. In *Social information retrieval systems: Emerging technologies and applications for searching the Web effectively*. IGI Global, 270–288.

[20] Brickley Dan and Guha R.V. 2014. RDF Schema. - W3C Recommendation. https://www.w3.org/TR/rdf-schema/. (2014). Last Accessed: 2018-08-16.

[21] Anastasia Dimou, Miel Vander Sande, Pieter Colpaert, Ruben Verborgh, Erik Mannens, and Rik Van de Walle. 2014. RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data.. In *LDOW*.

[22] Anastasia Dimou, Miel Vander Sande, Jason Slepicka, Pedro Szekely, Erik Mannens, Craig Knoblock, and Rik Van de Walle. 2014. Mapping hierarchical sources into RDF using the RML mapping language. In *Semantic Computing (ICSC), 2014 IEEE International Conference on*. IEEE, 151–158.

[23] W3C OWL Working Group. 2012. OWL 2 Web Ontology Language. - W3C Recommendation. https://www.w3.org/TR/owl2-overview/. (2012). Last Accessed: 2018-08-16.

[24] W3C OWL Working Group. 2013. SPARQL 1.1 Overview. - W3C Recommendation. https://www.w3.org/TR/sparql11-overview/. (2013). Last Accessed: 2018-08-16.

[25] Dick Hardt. 2012. *The OAuth 2.0 authorization framework*. Technical Report.

[26] Georgy Kopanitsa, C Hildebrand, Jürgen Stausberg, and KH Englmeier. 2013. Visualization of medical data based on EHR standards. *Methods of information in medicine* 52, 01 (2013), 43–50.

[27] Nicholas D Lane, Ye Xu, Hong Lu, Andrew T Campbell, Tanzeem Choudhury, and Shane B Eisenman. 2011. Exploiting social networks for large-scale human behavior modeling. *IEEE Pervasive Computing* 10, 4 (2011), 45–53.

[28] Luís Miguel Oliveira Machado, Renato Rocha Souza, and Maria da Graça Simões. 2018. Semantic Web or Web of Data? A diachronic study (1999 to 2017) of the publications of Tim Berners-Lee. (2018).

[29] Anmol Madan, Sai T Moturu, David Lazer, and Alex Sandy Pentland. 2010. Social sensing: obesity, unhealthy eating and exercise in face-to-face networks. In *Wireless Health 2010*. ACM, 104–110.

[30] M Scott Marshall, Richard Boyce, Helena F Deus, Jun Zhao, Egon L Willighagen, Matthias Samwald, Elgar Pichler, Janos Hajagos, Eric PrudâĂŹhommeaux, and Susie Stephens. 2012. Emerging practices for mapping and linking life sciences data using RDFâĂŤA case series. *Web Semantics: Science, Services and Agents on the World Wide Web* 14 (2012), 2–13.

[31] Kenta Matsumura and Takehiro Yamakoshi. 2013. iPhysioMeter: a new approach for measuring heart rate and normalized pulse volume using only a smartphone. *Behavior research methods* 45, 4 (2013), 1272–1278.

[32] Sai T Moturu, Inas Khayal, Nadav Aharony, Wei Pan, and Alex Pentland. 2011. Sleep, mood and sociability in a healthy population. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 5267–5270.

[33] K Thomas Pickard and Melanie Swan. 2014. Big desire to share big health data: A shift in consumer attitudes toward personal health information. In *2014 AAAI Spring Symposium Series*. 2168–7161.

[34] Roberto Reda, Filippo Piccinini, and Antonella Carbonaro. 2018. Towards Consistent Data Representation in the IoT Healthcare Landscape. In *Proceedings of the 2018 International Conference on Digital Health*. ACM, 5–10.

[35] Cyganiak Richard, Wood David, and Lanthaler Markus. 2014. RDF 1.1 Concepts and Abstract Syntax - W3C Recommendation. https://www.w3.org/TR/rdf11-concepts/. (2014). Last Accessed: 2018-08-16.

[36] Ishida Richard. 2008. An Introduction to Multilingual Web Addresses. https://www.w3.org/International/articles/idn-and-iri/. (2008). Last Accessed: 2018-08-16.

[37] Minna Ruckenstein and Mika Pantzar. 2017. Beyond the quantified self: Thematic exploration of a dataistic paradigm. *New Media & Society* 19, 3 (2017), 401–418.

[38] Manuel Salvadores, Paul R Alexander, Mark A Musen, and Natalya F Noy. 2013. BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF. *Semantic web* 4, 3 (2013), 277–284.

[39] Feichen Shen and Yugyung Lee. 2018. BioBroker: Knowledge Discovery Framework for Heterogeneous Biomedical Ontologies and Data. *Journal of Intelligent Learning Systems and Applications* 10, 01 (2018), 1.

[40] Latanya Sweeney, Akua Abu, and Julia Winn. 2013. Identifying participants in the personal genome project by name (a re-identification experiment). *arXiv preprint arXiv:1304.7605* (2013).

[41] Binyam Tilahun, Tomi Kauppinen, Carsten Keßler, and Fleur Fritz. 2014. Design and development of a linked open data-based health information representation and visualization system: potentials and preliminary evaluation. *JMIR medical informatics* 2, 2 (2014).

[42] Ben Vandervalk, Luke McCarthy, and Mark Wilkinson. 2008. CardioSHARE: Web Services for the Semantic Web. *Semantic Web Challenge* 2008 (2008).