



Heterogeneous self-tracked health and fitness data integration and sharing according to a linked open data approach

Roberto Reda¹ · Filippo Piccinini² · Giovanni Martinelli² · Antonella Carbonaro¹

Received: 2 October 2020 / Accepted: 21 July 2021 / Published online: 5 August 2021
© The Author(s) 2021

Abstract

The huge volume of data gathered from wearable fitness devices and wellness appliances, if effectively analysed and integrated, can be exploited to improve clinical decision making and to stimulate promising applications, as they can provide good measures of everyday patient behaviour and lifestyle. However, several obstacles currently limit the true exploitation of these opportunities. In particular, the healthcare landscape is characterised by a pervasive presence of data silos which prevent users and healthcare professionals from obtaining an overall view of the knowledge, mainly due to the lack of device interoperability and data representation format heterogeneity. This work focuses on current, important needs in self-tracked health data modelling, and summarises challenges and opportunities that will characterise the community in the upcoming years. The paper describes a virtually integrated approach using standard Web Semantic technologies and Linked Open Data to cope with heterogeneous health data integration. The proposed approach is verified using data collected from several IoT fitness vendors to form a standard context-aware resource graph, and linking other health ontologies and open projects. We developed a web portal for integrating, sharing and analysing through a customisable dashboard heterogeneous IoT health and fitness data. In this way, we are able to map information onto an integrated domain model by providing support for logical reasoning.

Keywords Health and fitness datasets · Linked open data · Semantic web · Ontology · Data integration

Mathematics Subject Classification 68-02

Roberto Reda and Filippo Piccinini have contributed equally to this work.

✉ Antonella Carbonaro
antonella.carbonaro@unibo.it

Extended author information available on the last page of the article

1 Introduction

Nowadays, wearable devices have significantly grown in popularity and recent statistics have shown that around 50% of people in developed countries make use of these devices to monitor fitness or physical activity [1]. Practically, people can keep track of their daily physical activities and constantly unobtrusively monitor their health status at an affordable price, or even at no cost at all in case of free mobile applications.

Commercial wearable devices include a large variety of apparatuses, ranging from sophisticated smart-watches to simple belt-mounted pedometers. The measuring capabilities include vital signs (i.e., blood pressure, heart rate, body temperature, oxygen saturation) and human body actions (i.e., steps taken, flights climbed, distance travelled) [2]. An important subset of wearable devices is the fitness trackers: cost-contained electronic bracelets with limited computation resources, although with the capability to keep track of the most important fitness-related measurements of the wearer, such as steps taken, distance walked, distance run, and heart rate [3, 4]. Users then have access to the collected data through smart-phone applications, Web portals or, if equipped, directly on device monitors.

However, the worldwide recorded data come from a variety of different heterogeneous sources and are represented with their own proprietary format depending on the device's manufacturer (Fig. 1a). This heterogeneity characterises all the Internet of Things (IoT) health and fitness datasets and, together with the typical huge volume of data, makes data sharing and integration extremely difficult. Accordingly, data heterogeneity is one of the main open challenges that need to be addressed to fully exploit the potential of the health data [5]. Section 2 of this work deals with this issue.

The process of gathering and integrating data from scattered IoT sources is normally done manually by researchers and domain experts [6]. This process is cumbersome, 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 and interpreting datasets from multiple distributed sources in a unified way in order to make them freely available to the research community (Fig. 1b). We propose to convert heterogeneous IoT raw data collected by a multitude of different devices into Resource Description Framework (RDF) graphs [7]. The homogenised datasets are then be stored in a structured format and exposed publicly via a SPARQL endpoint for accessing and querying. All the practical details are reported in Sect. 3.

Another open problem is health and fitness data robustness. Given the importance of the data to be collected, especially when used in medical applications such as post-traumatic rehabilitation and cancer prediction [8] (Fig. 1c), activity monitors must be precise and reliable. Accordingly, a great concern has been shown in the assessment of the validity and reliability of fitness trackers. It is difficult to evaluate and correctly consider the existing comparative studies of different devices due to the lack of standardised testing protocols and experimental

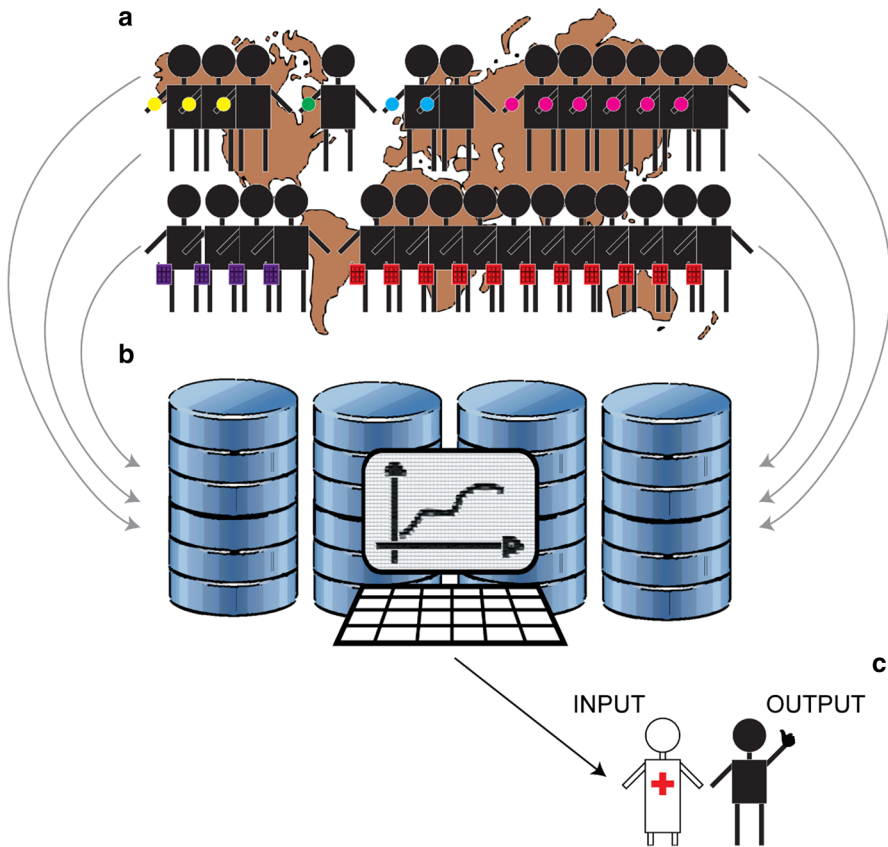


Fig. 1 **a** Health and fitness trackers **b** data integration to provide a comprehensive view of **c** patient's health

methodologies [9, 10]. An overview of the methodologies and gold standards to assess validity and accuracy of commercial activity monitors can act as guidance to clinicians and researchers willing to employ fitness trackers in their programmes and provide further resources and useful indications for future research on commercial wearable's validity assessment. Section 4 deals with this issue.

The validity and accuracy of the collected data is a fundamental issue. The heterogeneous nature of the collected information remains a problem limiting the community from really exploiting the opportunities provided by a large volume of health and fitness data. The number of research centres publicly disclosing health data is increasing but exploring datasets and extracting information is difficult and time-consuming for healthcare professionals. Although several attempts have been recently made to integrate different data, this topic is still quite new and remains unexplored [5].

In order to standardise data collection and integration and to allow users to achieve a common view of the available information, we designed and developed the

IoT Fitness Ontology (IFO). In Sect. 5 we introduced the IFO ontology, our domain representation model comprising the most common and important concepts related to the IoT fitness devices and wellness appliances. IFO allows to transform into RDF graphs the data stored in semi-structured formats by the multitude of heterogeneous IoT fitness sources.

Section 6 describes our approach using Linked Data and Semantic Web (SW) technologies to overcome this problem by modelling, sharing and interlinking IoT self-tracked health context information. The novelty of the proposed approach lies in exploiting SW and Linked Open Data (LOD) technologies to explicitly describe the meaning of the domain concepts and to facilitate interoperability and data integration, in order to construct a unified interlinked data model and enable semantic reasoning capabilities over it. We designed a LOD portal for the standardisation of the collection and integration of IoT health and fitness datasets which stems from our previous work [11]. In this work, we thoroughly described the development of the web platform and the design of the supporting ontology which lies at core of the process. The LOD portal may become a reference point for collecting, sharing and analysing IoT health and fitness data in structured format, accessible to domain experts, scientists and the web community without any restrictions by any form of patent or licensing. Differently from other platforms for sharing personal health data (PHD), such as Kaggle¹ or Open Humans² which redistribute users' data directly in raw formats (i.e., unstructured or semi-structured serialisation formats), a novel aspect of our portal consists in providing a semantic representation of the IoT datasets. Our IFO addresses the problem of data provided in heterogeneous formats by formally clarifying what the data describe, thus facilitating the integration and the analysis of the datasets, promoting innovative ways to reuse the data.

Finally, in Sect. 7, the conceptual architecture and sample data collected from several IoT fitness vendors are presented to formally encoding domain concepts and semantics of the collected data and to verify the described proposal. The obtained standard context-aware resource graph is linked to other health ontologies and open projects to map information onto a specialized domain model by providing support for logical reasoning.

2 Integrating health and fitness data

In the past years, the IoT industry has seen a proliferation of consumer devices for health and fitness tracking. The wearable technologies market alone is anticipated to grow from 325 million connected devices in 2016 to 929 million devices by 2021 [12]. The huge volume of data collected by these devices has enormous potential for the healthcare sector, especially combined with advanced Artificial Intelligence (AI) analytics techniques, for instance automated reasoning. From a data-centric perspective, the main issue that afflicts the IoT landscape is the presence of data silos caused

¹ <https://www.kaggle.com/>

² <https://www.openhumans.org/>.

by the heterogeneity of representation formats and the lack of interoperability. Such conditions prevent healthcare professionals from getting an integrated overview of health data (i.e., a representation of the complete knowledge) and an efficient data analysis process.

SW technologies 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. The SW architecture is based on a layered approach, and each layer provides a set of specific functionalities. Semantic layers include ontology languages, rule languages, query languages, logic, reasoning mechanisms, and trust. Ontologies constitute the backbone of the SW expressing concepts and relationships of a given domain, and specify complex constraints on the types of resources and their properties. A more detailed introduction to the major themes in SW research and data representation in the IoT healthcare domain can be found in [13]. Interoperability is particularly relevant in the IoT self-tracked health data domain, where a multitude of diverse vendors collect the same type of data but store and exchange them in many different ways. Semantics gives a structure to data and captures the meaning. In recent years, there has been a great deal of interest in the development of semantic-based systems to facilitate data integration and knowledge representation of heterogeneous data. Within the SW context, ontologies play a key role in resource representation, since they explicitly define concepts and relationships related to a particular domain in a structured and formal way (i.e., ontologies are machine-processable) [13, 14, 15]. For example, Alamri recently proposed [16] a semantic interoperation middleware for IoT data in electronic healthcare records (EHR) data domain to improve patient healthcare, enabling health providers to monitor their patients outside the clinic. He specifies complex constraints on the types of resources and allows expressiveness and powerful logical inferences. The existing fitness and wellness data aggregators rarely make use of SW technologies. Therefore, they partially solve data integration, sharing, and analysis problems.

In a recent publication, we listed the main solutions today available for collecting and integrating health and fitness data [13]. Briefly, Apple Health³ is an information hub for integrating in a single location point data from eHealth apps for iOS devices. Apple HealthKit⁴ provides APIs that allow third-party developers and medical sensor manufacturers to directly store their data within the Apple Health app. Apple allows users to store and aggregate health content which can optionally be exported in XML format, or encrypted and uploaded onto Apple's iCloud servers. On the other hand, apps and devices that rely on HealthKit are restricted to run on iOS platforms only. Google Fit⁵ is the Apple Health equivalent for Android operating systems. It is currently limited to fitness data only, whilst Apple Health supports a wider variety of medical data. Google Fit aggregated content is accessible via the Web portal or through a REpresentational State Transfer (REST) APIs. Google Fit defines fixed sets of data types which can be

³ <https://www.apple.com/lae/ios/health/>.

⁴ <https://developer.apple.com/healthkit/>.

⁵ <https://www.google.com/fit/>.

stored, and third-party developers need to inform Google to add and share new ones. Google Fit and Apple Health are intended to be data aggregators for their respective ecosystems and let health and fitness applications, as well as wearable devices, gather health information in one single location point. However, they are not interoperable among each other or with other systems; therefore, data remain confined to their respective platforms. MyFitnessCompanion [17] is a health and fitness app which aims to enable users to aggregate their data in one place in a similar way to Apple Health and Google Fit. It integrates off-the-shelf a significant number of commercially available devices; it can interact with a wide range of wireless devices and wearable health trackers and can also aggregate data from third-party apps. However, MyFitnessCompanion can be used only on Android platforms. MELLO [18] is an ontology for representing health-related and life-logging data including definitions, synonyms, and semantic relationships. The unified representation of lifelog terms facilitated by MELLO can help to describe an individual's lifestyle and environmental factors, which can be included with user-generated data for clinical research and thereby enhance data integration and sharing, although, a SW system needs a mapping process to semantically annotated values within the data sources according to an ontology. Recently, Patel et al. created SWoTSuite [19], which is an infrastructure that enables SWoT (Semantic Web and Internet of Things) applications. It takes high-level specifications as input, parses them and generates code that can be deployed on IoT sensors at the physical layer and IoT actuators, and user interface devices at the application layer. SWoTSuite hides the use of SW technologies as much as possible to avoid the need for designing ontologies, annotating sensors data, and using reasoning mechanisms to enrich data.

In order to (a) accumulate and represent knowledge in a wide range of different databases, services and vocabularies, and (b) obtain an integrated view of health data and a representation of the complete knowledge, we propose a virtual integration approach using a distributed architecture based on remote sources and their access interfaces, without creating a single physical knowledge representation. Efforts are focused on the process of interlinking data silos given that the entities from distributed sources usually complement each other. The knowledge then becomes a logical structure represented by mapping remote data sources using ontologies to create and maintain relevant *owl:sameAs* links. For example, *owl:sameAs* semantic links connect individuals from DBpedia, PubMed and Mesh source graphs using their URIs. Existing vocabularies with respective URIs are reused to eliminate the need to introduce a custom URI naming. Table 1 summarises the main technical features of the physical and virtual approaches for knowledge representation [20, 21]. The physical integration of data sources requires the handling of certain issues. Custom URIs need to maintain consistent alignment between remote data sources and a local schema in the new unified local storage graph. Moreover, physical integration requires sophisticated data fusion mechanisms. Conversely, the logical structure obtained from a virtual integration outperforms separated physical sources in terms of data scalability and interlinking.

Table 1 Main features of physical and virtual health and fitness data integration approaches

Criteria	Physical approach	Virtual approach
Description	Implies the transformation of data sources into a single unified local storage	Implies the use of schema mapping to maintain remote data sources in their original status and formats
Resulting knowledge graph	A physically new knowledge graph where one data source is chosen as the primary vocabulary and the other data sources must be converted	A logical structure that allows mapping between several remote data sources. The resulting graph comprises the original datasets and the set of semantic links between their entities
Naming	The new graph requires custom URIs	URIs are reused so that there is no need to introduce custom naming
Maintenance	A changed fact requires new transformation of data source into target unified local storage	A changed fact requires the adjustment of involved semantic links. The knowledge graph is updated naturally as soon as the fact is processed by the linking engine
Cost	Maintaining data fusion between graphs has a high computational cost	Creating schema mapping has a high computational cost

3 Health and fitness open-access data

The enormous amount of self-tracked health information collected by users through smart fitness devices offers important opportunities to the research community. The market for these devices has been growing steadily over the last few years and continues to do so. The ecosystem of mobile health solution is very complex and the need to provide integrated and open access data is strong. The mHealth Developer Economics is a global research program analysing the digital health and mobile health market since 2010. The last mHealth Developer Economics survey cycle (<https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/>) describes this development. Surveys from participants have been collected globally, with most of the answers coming from Europe (47%) and North America (36%). The other participants are from Asia-Pacific (11%), South America (4%) and Africa (2%). This year there are 325,000 health and fitness and medical apps available on all major app stores and 78,000 new health apps have been introduced since last year. However, only 27% of all mobile health app publishers have already opened their apps for others by directly offering access to a wealth of valuable data for instance through an API. 42% of mobile health apps connect to sensors and wearables. Fitbit is the most connected-to sensor/wearable (52%), followed by iHealth⁶ and Withings.⁷ Smartwatches are becoming more attractive for mobile health app publishers, replacing other wearables. There were more than 50 Wear OS watches available in Q2/2018 from a range of third-party manufacturers like LG,⁸ Fossil,⁹ Ticwatch,¹⁰ Asus,¹¹ and Huawei.¹² In this fragmented situation, it becomes important to furnish an integrated approach. API aggregation services bring together APIs from different sources into one single hub, pulling data from different sources, combining it and making it available for third parties.

Apple HealthKit is by far the most popular service with two thirds (63%) of API users opting for Apple. Number two is Google Fit (45%). All other API service providers are used by 20% or less: Open mHealth, Samsung Health, Human API, Validic and Qualcomm Life. HealthKit, like other Apple products, is restricted to run on iOS devices. Google has a strict policy regarding what data developers can share via Google Fit. Google's policy is that health data cannot be published [22]. Samsung Health app developers can use an existing set of data types and can extend this set with their own data types. But one disadvantage of this approach is the vendor lock-in, which means further technology-driven innovations become difficult due to the vendor-specific interconnections among the different parts of

⁶ <https://ihealthlabs.com>.

⁷ <https://wethings.com>.

⁸ <https://www.lg.com>.

⁹ <https://www.fossil.com>.

¹⁰ <https://www.ticwatch.com>.

¹¹ <https://www.asus.com>.

¹² <https://www.huawei.com>.

the architecture. Open Humans is an open-source project which aims to make more health-related data available for scientists. The online portal allows users to upload, store and share their personal data such as genetic, social media, activity and health data gathered through IoT devices. Open Humans is a coproduced model of data community, where users are increasingly encouraged and facilitated to improve their healthcare information dataset for clinical and research purposes. Another well-known example of data community is Kaggle: it provides a platform to store and share a variety of dataset formats. Kaggle's core idea is to facilitate the analysis of data by allowing outsiders to model it. To do that, the company organises competitions in which anyone interested can participate. Kaggle adopts a crowd-sourcing approach to collect datasets from companies, scientists and users, including IoT wearable data. The dataset repository listing can be viewed by size, file type, most votes and tags. PhysioBank¹³ is a different solution to store and share health datasets. The project collects databases of physiologic signals and offers free access to the research community since 1980 via Web. Successfully read and manipulated, the databases require specialised software: the distributed toolkits supply methods for reading and writing signals and annotations in many formats and can be linked to user-written applications in C, C++, and Fortran. DataGraft project [23] provides a set of tools and methodologies for open-data transformation and hosting services. DataGraft is designed to be scalable and reliable in a cloud-based environment. DataGraft's features include RDF data publication and querying. It was developed to provide easy-to-use tools for users who consider the existing approaches to data transformation, hosting, and access too costly and/or technically complex. ResearchKit¹⁴ is an open-source framework introduced by Apple that allows the use of health data directly from users' smartphones. ResearchKit collects medically relevant data obtained using the built-in capacity of the mobile device and secure data in a central repository, in compliance with regulatory requirements. The mobile application can communicate with connected devices to collect data via additional sensors, such as a heart-rate monitor on a watch or fitness band.

All these platforms are important initiatives to publish and share IoT health datasets online, but richer semantics of data are needed to resolve the heterogeneity problem and allow information integration and reuse. A notable example is Bio2RDF [24]. Bio2RDF addresses the data integration problem by integrating publicly available databases in bioinformatics. Bio2RDF uses SW technologies to create a knowledge space of RDF documents linked together, sharing a common ontology. Bio2RDF scripts convert heterogeneously formatted data into RDF common format, without an attempt to marshal data into a single global schema; Bio2RDF currently provides the largest network of Linked Data for Life Sciences. BioPortal [25] is an open repository of biomedical ontologies that allows multiple mechanisms for content updates, provides access via Web services and provides support to integrate data from a variety of biomedical resources. BioPortal users can browse, search and visualise ontologies. The Web interface supports the evaluation and evolution of

¹³ <https://www.physionet.org/physiobank>.

¹⁴ <https://developer.apple.com/researchkit>.

ontology content by providing features to add notes to ontology terms, mappings between terms and ontology reviews.

While all these projects share the common elements of longitudinal integration of heterogeneous relevant data, in some cases, even in health-related fitness data, each of them focuses on a relatively narrow set of measurements, or relies on commercial or custom data storage and analysis architectures that do not provide a unifying model to promote open data sharing and analysis from multiple sources.

4 Health and fitness data robustness

Another aspect that must be considered when analysing a large amount of data from IoT self-tracked devices is their reliability, especially when the data are then used in medical applications. It is worth noticing that significant measurement inaccuracies not only jeopardise the analyses, but can even result in harmful consequences for both consumers and clinician users in case of a wrong diagnosis. For example, an underestimated physical activity may lead athletes to overcompensate by over-exercising, then risking exhaustion or injuries. On the other hand, an overestimated total energy expenditure could lead patients to reduce their activity levels and adherence to activity prescriptions [26]. Furthermore, following the common myth that lifestyle modifications are promising strategies to reduce cancer risk [27], several studies have been performed in the last years from a number of Nations to infer scientific connections between people's lifestyle and tumour incidence [8, 29]. The huge amount of data stored comported (and will comport) significant costs for the supporting research Institutions, but today we have several scientific articles with discordant conclusions, probably due to data acquired without using standardised approaches and devices tested in different scenarios. More in-depth testbeds for devices should be performed, especially when proposing large population screenings. For example, results of tests for steps-tracking conducted in laboratory settings (e.g., on a treadmill) may significantly differ from results of tests conducted in everyday conditions [29]. In fact, the accuracy of consumer wearable devices is affected by a variety of different factors such as the wearer's anthropometric characteristics, health conditions and the kind of activity which is being performed. Devices tested on healthy individuals, which report with a high accuracy, may not be equally suitable for monitoring patients with gait impairments [23]. Moreover, it should be noted that other important factors affect the validity of wearable devices and the precision of the experimental tests, such as sensor positioning, battery status, firmware version, device configuration and temporal granularity of the acquired data. These aspects are often overlooked or not reported, thus making the reproducibility of the experiments impossible and the result data not comparable across various studies [10]. Since everyday activities require complex body movements characterised by fluctuations in direction, speed, intensity and body parts involvement, it is clearly impossible for a fitness tracker worn on a single limb to capture all of them. Thus, activity monitors are always subject to measurement errors and activity misclassifications. For instance, in the case of wrist-worn devices, activities that require high levels of wrist action, such as washing hands, may result in the detection of

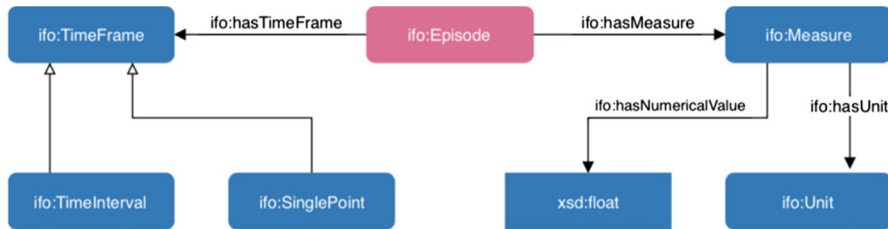


Fig. 2 Excerpt of the IFO ontology. An Episode is any event that can be recorded by an IoT fitness device and it constitutes the fundamental abstraction mechanism of the IFO ontology. Episodes are always precisely collocated in time and can be numerical quantified

increased activity level [30]. Similarly, users' physical conditions can impact the measurement accuracy as well, for instance patients with movement disorders, such as Parkinson's disease, may obtain an overestimation of the overall physical activity, whilst a person with limited arm movement may see an underestimation. Moreover, improper placement of the devices is another important factor that can significantly affect the ability of the device to correctly assess the actual activity, thus compromising the measurements [30]. In this context, it would be important to refer to an overview of the existing literature and a discussion of the methodologies and standards used in scientific studies to assess the validity and accuracy of the used devices.

5 The IFO ontology: design, process and implementation

The IoT Fitness Ontology (IFO)¹⁵ is a domain ontology which aims to represent the most common and important concepts within the domain of the IoT fitness devices and wellness appliances. The list of products and vendors that were taken in consideration during the design process includes: Apple Health, Microsoft HealthVault, Google Fit, Fitbit, Jawbone, Strava, Runtastic, iHealth and Nokia Health. The key terms used in the ontology are the nouns describing generic types of physical activities and physiological parameters with no relation to specific brands. Examples of terms used about physical activities are: Steps, Running, Walking, Swimming, ActivityIntensity, FlightsClimbed. Examples of terms used about physiological parameters are: HeartRate, BodyTemperature, BodyWeight, BloodPressure, CaloriesBurned. Examples of other general terms are: Meditation, TemporalRelationship, BodyPosture, Measure, Statistics, TimeFrame, MassUnit.

Within the IFO ontology we organised the classes representing the concepts in a classic hierarchical fashion in a top- down approach. The ontology is built around the root class Episode which represent the set of all possible events that can be measured by an IoT wellness device. For example, an episode could be the heart beat rate measured during a running training session by a wearable wrist worn heart rate monitor or the body weight of the user measured by a smart scale. To each

¹⁵ <http://purl.org/ifo>.

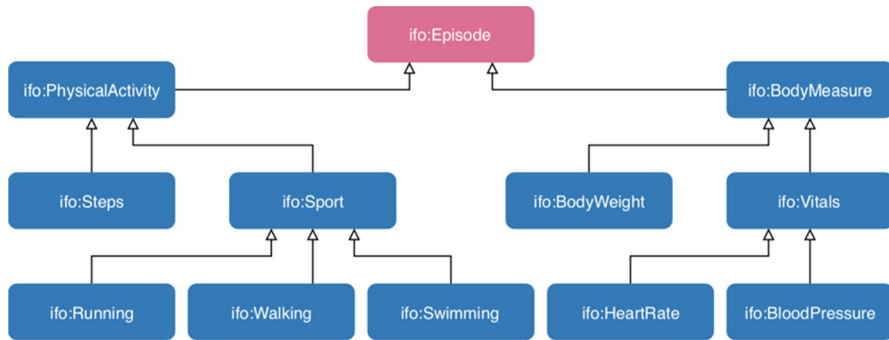


Fig. 3 Excerpt of IFO ontology hierarchy. Episodes are grouped into two main categories: physical activities and body measurements. Physical activities are the kind of events which involve a body movement (e.g., a walk) and are typically measured by wearable devices. Body measurements regard the physiological readings normally collected using health appliances (e.g., smart scales, digital blood pressure meters)



Fig. 4 Excerpt of the IFO ontology. Episodes can be augmented with metadata such as individual's personal information or geolocation position

episode is always associated a time reference and a numeric measurement value with the related unit of measurement. The time reference can be a single point in time or a time interval, that is, the start time and the end time of the event. This information are essential because they allow to numerical quantify the object of the event and give it a temporal collocation and duration (Fig. 2).

Two main categories of episodes can be distinguished: (a) the physical activities; (b) the body measurements (Fig. 3).

Physical activities encompass any kind of activity involving body movement such as walking, running, swimming or steps taken. Body measurements, on the other hand, are relative to the physiological parameters of a person such as the body weight or body height or the person's vital signs such as the heart rate or the blood pressure. Minor categories of episodes that the IFO ontology defines, concern the sleep and the meditation. It is noteworthy to underline that some measurements require more than a single numerical value such as the blood pressure. The blood pressure is measured in millimetres of mercury (mmHg) and is written as two numbers (e.g., 120/80 mmHg). The first (120 in the example aforementioned) number is the systolic blood pressure, and the second number

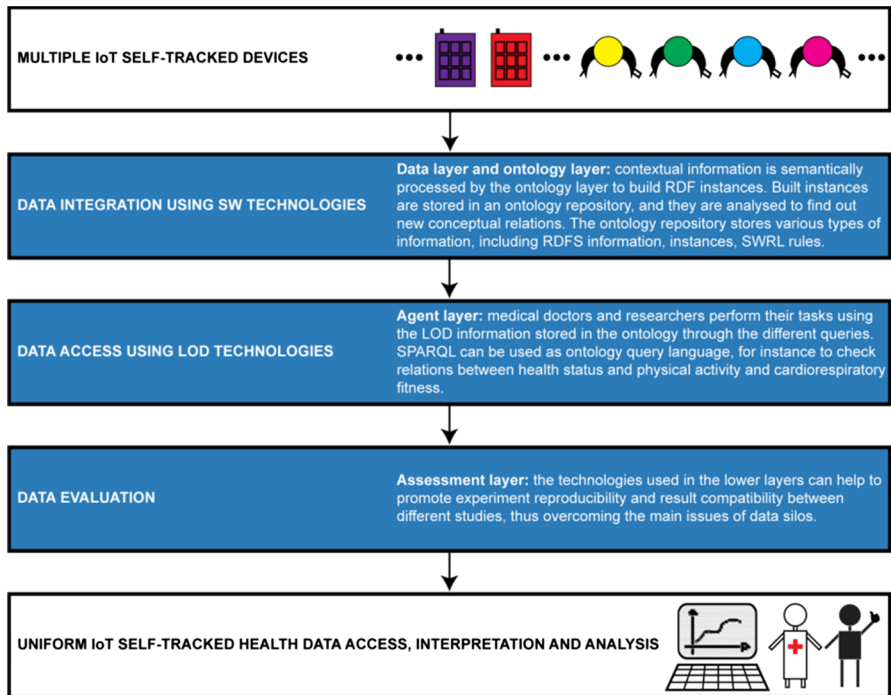


Fig. 5 From different IoT self-tracked data to a uniform model. Three-layered conceptual architecture to transform multiple IoT self-tracked data in a uniform model to access, model and analyse data. The enriched data model allows its reuse and a logical reasoning on knowledge representation

(80) is the diastolic blood pressure. Systolic blood pressure and diastolic blood pressure according to the IFO ontology are two separated episodes. Other components of the IFO ontology are the OWL class *Measure* and the class *Time-Frame* which they respectively model the measurement and the time reference; these two classes are associate to the *Episode* class through the OWL properties *hasMeasurement* and *hasTime Frame* as shown in Fig. 2. Metadata such as geolocation coordinates or individual's information can be optionally added to episodes (Fig. 4).

Devices used to acquire data about an episode are represented in the IFO ontology by the class *InputSource* and are classified in *Wearable* for wearable devices, *Appliance* generic systems, *Smartphone* for mobile applications and *UserTyped* for episodes recorded manually by the user.

Object properties have been defined to model the relationships among concepts. The two most important object properties relate an episode to its measure (i.e., *hasMeasure*) and to its time reference (*hasTimeFrame*). Units of measurement were modelled as OWL individuals since are concepts that cannot be specialised anymore in the hierarchy. To achieve a better integration with other systems and better specify the meaning of each class, references to other

standardised ontologies such as SNOMED-CT were made. Personal information (e.g., date of birth) was based on FOAF ontology and the Basic geo (WGS84 lat/long) vocabulary was used for the geospatial locations.

6 A layered approach to integrate iot health and fitness data

In the previous Sections we introduced how the use of semantic technologies and LOD can provide standardized frameworks for the concept representation in health and fitness landscape. IoT low-level data can thus be transformed into an enriched information model that allows its reuse and a logical reasoning on the knowledge representation. The conceptual architecture of such a system is illustrated in Fig. 5, comprising of three components, namely (i) data integration layer, (ii) data access layer and (iii) data evaluation layer. (i) The data integration layer collects domain datasets from users or remote servers and transforms semi-structured format input data into an RDF graph. Datasets are semantically annotated according to reference ontologies and stored within a triple store server. (ii) The data access layer controls data access and bridges the clients with the system via service protocols, allowing users to interact with the system using the Web-based access or the SPARQL endpoint. (iii) The data evaluation layer helps to promote the reproducibility of the experiments and the comparability of the results across various studies, overcoming the main issues of data silos. The novelty of the proposed approach lies in exploiting SW and LOD technologies to explicitly describe the meaning of the domain concepts and to facilitate interoperability and data integration, to construct a unified interlinked data model and enable semantic reasoning capabilities over it. We suggest the use of SW technologies and LOD in order to ensure standardized frameworks and to facilitate reuse and consumption of these data. Accordingly, we propose the three-layered architecture summarised in Fig. 5 to implement semantic interoperable systems, addressing all four types of interoperability (system, structural, syntactic and semantic) in order to exchange information without the loss of meaning or intent.

To further improve the usability of semantically integrated data, starting from the early work described in [11], we developed 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 by Marshall et al. in [31].

We developed the web portal using JavaServer Pages (JSP)¹⁶ (JavaServer Pages Technology, n.d.) as back-end technology. The experimental web portal is available at: <http://137.204.74.19:8080/IFOPlatform/welcomePage.jsp>. On the Help page of the web portal a video tutorial and some sample datasets for testing purposes are provided. Precisely, 4 sample datasets in different file formats are available: (I)

¹⁶ <http://www.oracle.com/technetwork/java/index-jsp-138231.html>.

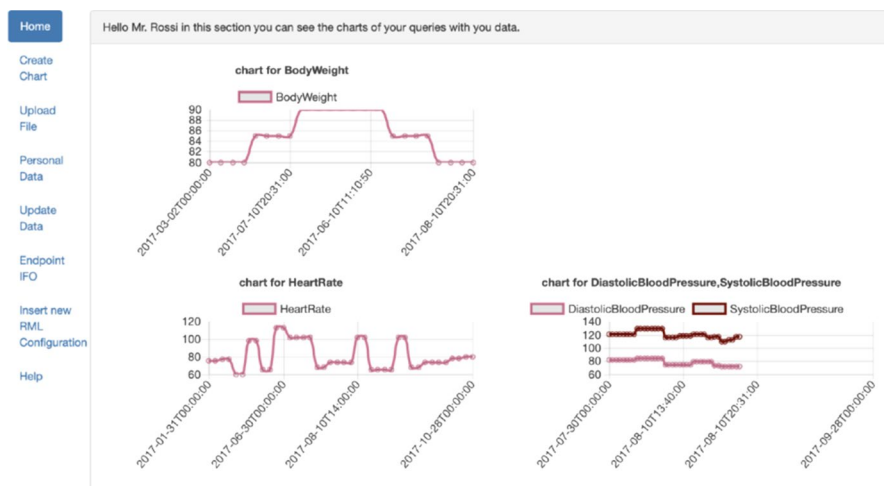


Fig. 6 A screen shot of the web-based dashboard showing some vital signs data charts and a customised body weight chart generated through a user defined SPARQL query

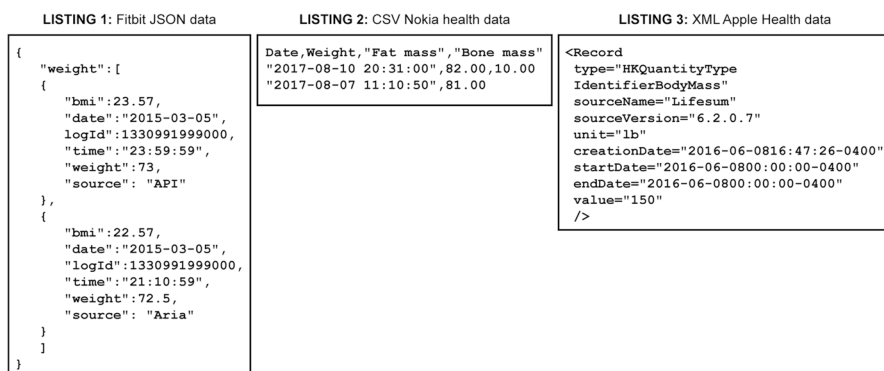


Fig. 7 Different data formats. *Listing 1*: An excerpt of body weight data retrieved in JSON format using Fitbit proprietary APIs. *Listing 2*: An excerpt of body weight data collected by a Nokia Health smart scale in CSV format. *Listing 3*: An excerpt of health data manually exported in XML format from the Apple Health app

“blood_pressure.csv”, file format: Nokia Health; (2) “blood_pressure.xml”, file format: Health Kit; (3) “fitbitWeight.json”, file format: Fitbit; (4) “weight.csv”, file format: Nokia Health. Furthermore, we also provided a video tutorial¹⁷ to guide step-by-step the users in exploiting the different functionalities of the portal. Briefly, our LOD portal is capable of: (a) collecting IoT fitness data manually entered by users or automatically retrieved from remote repositories; (b) integrating and storing IoT datasets semantically annotated according to a reference ontology; (c) visualising

¹⁷ http://isi-idata.csr.unibo.it:8080/IFOPPlatform/video_tutorial.mp4.

information through a customisable dashboard; (d) sharing datasets adhering to LD principles.

The portal allows users access their PHD through a customisable dashboard which can provide multiple views of the integrated datasets. RDF graphs offer unique opportunities since they enable to bind data to visualisations in unforeseen and dynamic ways. 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 [32]. To take advantage of the flexibility provided by RDF graphs, we made the dashboard highly customisable by letting expert users to define the information to be displayed on charts through custom-made SPARQL queries (Fig. 6). Since disjoint-domains data integration is a key feature of SW, federated queries are also possible within the embedded dashboard. However, writing SPARQL queries is a challenging task for nontechnical users. For this reason, several preset queries for visualising common information (such as the heart rate or the blood pressure readings) in the form of time series are available on the dashboard by default.

7 Resulting integrated domain model

To demonstrate the usefulness of our conceptual architecture, we used data collected from several IoT fitness vendors. For example, the JSON code shown in Fig. 7 *Listing 1* is the response obtained after being authenticated and authorised to the Fitbit server and executing an HTTP GET request using Fitbit proprietary APIs. Figure 7 *Listing 2* shows “body weight” data collected by Nokia Health smart scale in CSV format. Figure 7 *Listing 3* shows an excerpt of data manually exported from Apple Health7 in XML format. Since for each vendor to obtain the required data defines its own specific proprietary API interface, deploying ad-hoc platforms could be very expensive and ineffective on the long run. These are the technologies that do not allow data to be distributed in LOD mode for possible sharing and reusing. Data access and data evaluation layers are needed to extract data and convert them into a standard RDF format, so to be distributed in open mode, that means queryable triplestores, for instance using RDF/XML or JSON-LD format.

For the mapping process we employed the RDF Mapping language (RML) [33] for mapping specification and the RML Processor for its execution. RML is a declarative source-independent mapping language which allowed us to express customised mapping rules for converting heterogeneous resources into RDF graphs according to a reference ontology (i.e., the IFO ontology). RML extends the W3C standard R2RML (R2RML: RDB to RDF Mapping Language, n.d.) which can define customised mappings only from data stored in relational databases. RML keeps the mapping definitions as in R2RML but encompasses broader variety of data format as input sources. Additionally, RML provides the vocabulary for defining the iterator pattern over the input data which allows us to explicitly specify how the source data that have to be accessed. Iterator patterns make use of target-specific query languages. For instance, an XPath expression can be used to specify an iterator over an

LISTING 1: Fitbit JSON data

```

<#FitbitBodyMass>

rml:logicalSource [
  rml:source "fitbitWeight.json";
  rml:referenceFormulation ql:JSONPath;
  rml:iterator "$.weight";
];

rr:subjectMap [
  rr:termType rr:BlankNode;
  rr:class ifo:Measure;
];

rr:predicateObjectMap [
  rr:predicate ifo:hasNumericalValue;
  rr:objectMap [
    rml:reference "@.weight";
    rr:datatype xsd:float;
  ];
];

```

LISTING 2: RDF graph

```

_:kWRuix2ft9 a ifo:BodyWeight ;
ifo:hasMeasure _:fxbMJQzZG8 ;
ifo:hasTimeInterval _:CrHFdBYBD8 .

_:fxbMJQzZG8 a ifo:Measure;
ifo:hasNumericalValue "73"
^^xsd:float;
ifo:hasUnit ifo:kg .

_:CrHFdBYBD8 a ifo:TimeInterval ;
ifo:endDate "2015-03-05 23:59:59"
^^xsd:dateTime ;
ifo:startDate "2015-03-05 23:59:59"
^^xsd:dateTime .

```

Fig. 8 Listing 1: An example of RML triples map which can be used to generate an RDF graph starting from a JSON file about body weight data collected by a Fitbit IoT smart scale. Listing 2: RDF graph representing IoT health data annotated according to the IFO ontology

XML document while a JSONPath expression can be defined in a similar way for a JSON document. RML as a source-independent mapping language is particularly useful within the IoT fitness context because different IoT devices vendors employ different data serialisation formats to represent and store information about the same concept. RML mapping specifications are based on one or more Triples Maps which define how the triples (i.e., the resulting RDF graph) are generated. A triple map contains a rule to generate zero or more RDF triples which share the same subject for each extract of data from the input source. A single triples map is composed by the Logical Source, the Subject Map and zero or more Predicate-Object Maps.

As an example, Fig. 8 Listing 1 shows a set of RML triples maps which can be used for generating an RDF graph starting from the Fitbit data about the body weight as proposed in Fig. 7 Listing 1. The logical source consists of the reference to the input source to be mapped, in this case the `fitbitWeight.json` file. The Reference Formulation, pinpoint by `rml:referenceFormulation`, specifies how references to the data occurs and, since RML uses references relevant to the input source, in this case JSONPath is used. The iterator specifies how to iterate over the input data, here is specified by the JSONPath expression: `$.weight`. The subject map consists of the template that defines the URI pattern used to generate the subject of the triple and optionally its type. A blank node is generated and the triple is typed as `fo:Measure`; `fo` is the name space used for the IFO ontology. A Predicate Object Map consists of a Predicate Map that specifies the predicate of the triple and an Object Map which specifies the object (one or more) of the triple. A JSONPath expression is used to point to the body weight value in the source `rml:reference"@.weight"`. The resulting RDF graph is shown in Fig. 8 Listing 2.

RML significantly 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. Several mapping specifications, manually written, 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 triple store.

Thanks to these layers, data from multiple heterogeneous sources are now exposed in a structured format and the different representations converted to standard RDF formats can be merged to form a context-aware resource graph. Conversely, ad hoc Web or desktop applications are able to synchronize body weight measurements between platforms with online clouds or local files by selecting source and target formats. These applications generally use their own file format and the availability of units and types of measurement depends on the specific supplier (e.g. www.weemple.com/weighthub). Now, the resource graph can be further expanded to include concepts, relationships and data from other ontologies and linked open projects. For example, we can add a triple: “*schema:angina_pectoris owl:sameAs snomedct: 194,828,000*” to indicate that the concept “*angina pectoris*” in the *Schema.org* vocabulary has the same meaning as that of “*snomedct:194,828,000*”, which is the ischemic heart disease concept under the SNOMED-CT clinical ontology, also related to the class: “*Angina co-occurrent and due to coronary arteriosclerosis*” and the subclass: “*Preinfarction syndrome and related to Family history: Angina in first degree female relative less than 65 years*”. Similarly, angina concept can be related to *DBpedia* (the central interlinking hub of the Web of Data containing millions of RDF links to other Web data sources), *Mesh* (Medical Subject Headings, a comprehensive controlled vocabulary for the purpose of indexing life sciences journal articles and books), or *Wordnet* (a fairly large on-line lexical reference system offering broad coverage of general lexical English relations) by adding the following triples: “*schema:angina_pectoris owl:sameAs dbpedia:552,599*”, “*schema:angina_pectoris owl:sameAs meshId:D000787*”, “*schema:angina_pectoris owl:sameAs wordnet:14,197,107*”, and “*schema:angina_pectoris owl:sameAs wordnet:14,131,521*”, which represent “*disease of the throat or fauces marked by spasmodic attacks of intense suffocative pain*” and “*the heart condition marked by paroxysms of chest pain due to reduced oxygen to the heart*” concepts. More details on a WordNet-based knowledge representation used for concept recognition and reasoning processes and on the use of ontologies to recognize the concepts in the domain ontology are described by Riccucci et al. [34].

We expanded the resource graph by linking it with other health ontologies, LOD, linked open health data and hierarchy concept graphs containing hyponyms and hypernyms. Accordingly, they can be accessed and queried in a uniform way using standard languages. Data visualisation in a personalised manner is now possible through a Web dashboard. The homogenised data are now available for statistical analysis, for example to monitor the percentage of individuals aged 15 or above who had cancer, or are overweight and obese, and also to survey health conditions and recourse to health services. Once the data is represented as RDF and exposed through a SPARQL endpoint we can combine them with other data belonging to a different LOD portal, because the different storage modalities are irrelevant from a SPARQL query perspective. A federated query “on the fly”, like the one aforementioned, without the support of

SW technologies, would have been complex to be formulated and executed. For instance, we could carry out a selective survey of risk factors affecting the health status of all families living in a specific area by cross-checking data extracted from devices with those from the registry of families in the territory. The interpretation and the processing of the resulted knowledge can be furthermore additionally enhanced owing to the SWRL to propose rules for different goals, such as to verify the proper functioning of the connected objects and the validity of the detected data and to provide the adequate service for patients. Summarizing, our proposed integration architecture combines the advantages of formally representing domain concepts, collecting data and their relationships allowing to map information onto a specialized domain model by providing support for logical reasoning [35–38].

The presented architecture allows acting at different levels on the validity of the functionality and the verification of the data collected by the sensors. As example, consider that the sensors are characterized by numerous properties such as data range, frequency, date and time, etc. We can set SWRL rules to automatically verify the values of vital signs detected by the sensors, with respect to the interval that they must assume to validate the operation of the sensor. The following rule performs the validity of sensor vital signs:

Rule Validity_vital_signs: $\text{IoT}(\text{?o}) \wedge \text{Sensing-device}(\text{?s}) \wedge \text{contains}(\text{?o}, \text{?s}) \wedge \text{Measurement}(\text{?m}) \wedge \text{detects}(\text{?o}, \text{?m}) \wedge \text{hasvalue}(\text{?m}, \text{?v}) \wedge \text{hasmaxValue}(\text{?o}, \text{?maxv}) \wedge \text{hasminValue}(\text{?o}, \text{?minv}) \wedge \text{swrlb:greaterThanOrEqual}(\text{?v}, \text{?minv}) \wedge \text{swrlb:lessThanOrEqual}(\text{?v}, \text{?maxv}) \rightarrow \text{validity}(\text{?m}, \text{true})$

Another SWRL rule can be set to determine the risk exposure of patients. For example, the following rule aims to verify the risk of a patient suffering from obesity (an event detected by the execution of another rule and passed as an input to the current rule) to be subject to cardiac attack if at the same time subject to events that highlight cardiac problems.

Rule cardiac_attack_risk: $\text{Patient}(\text{?p}) \wedge \text{has-event}(\text{?p}, \text{Obesity}) \rightarrow \text{has-risk}(\text{?p}, \text{Cardiac-failure})$

There are a lot of health situations in which IoT low-level data should be complemented by social and Web data, collective intelligence and domain ontologies (i.e., medical knowledge). These situations provide conditions for the establishment of a more complex scenario, but also more realistic one where, for example, heart rate, pressure and body weight are to be considered in conjunction with medical diabetes treatment as exenatide and liraglutide [39]. Our workflow enables data integration using semantic representation, reasoning technologies and incorporating domain knowledge into the computation. We established a broad workflow that takes a stream of low-level-encoded IoT information instances, transforms them to domain-level Ontology Web Language concepts [40], and reasons with them to generate knowledge.

8 Conclusions

The huge amount of self-tracked health information collected by users through smart IoT devices offers important opportunities to the research community. However, an effective and efficient exploitation of these data requires methods for accessing, integrating and analysing datasets from multiple distributed sources in a unified way. In this paper we focused on current, important needs in IoT self-tracked health data modelling and described an approach for the representation and the context-aware integration of IoT health and fitness data. The main goal is to understand current limits and future opportunities related to the large amount of open-access, robust and accurate IoT self-tracked health data. To address these opportunities the paper presented a virtually integrated approach using SW and LOD technologies. The proposed approach is verified using data collected from several IoT fitness vendors to form a standard context-aware resource graph, and linking other health ontologies and open projects. The paper showed how to map information onto an integrated domain model by providing support for logical reasoning. Such representations can be a viable and comprehensive solution for describing and integrating the heterogeneous IoT health and fitness data, thus overcoming the main issues of data silos.

These findings regarding obstacles, benefits, and facilitators can guide the development of smart systems and help researchers determine best practices when developing horizontally integrated schemes and harnessed by knowledge acquisition and capabilities sharing to fully exploit the potential of the IoT health and fitness devices. When these challenges will find positive proposals by the research community, we will see an explosion of knowledge in several fields, particularly in medical applications. For example, in Oncology, it will be possible to statistically analyse different cancer types and the relations between lifestyle and cancer incidence and see if sedentary habits really increase the cancer risk whilst regular physical activity and high cardiorespiratory fitness really have the opposite effect.

Acknowledgements We would like to thank Mattia Cantagalli (Lupo Team Full Contact, Amateur Sports Association, Faenza, Italy), Roberto Montevercchi (A.U.S.L. Romagna, Ravenna Hospital, Italy), Roberto Vespignani and Americo Colamartini (IRST IRCCS, Meldola, Italy) for sharing with us important opinions about fitness trackers; Richárd Tasi (Szeged, Hungary) for English revision of the manuscript.

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement. G.M. is supported by Horizon 2020—ONCORELIEF EU Project (ref: H2020-875392). This research received no other external funding.

Declarations

Conflicts of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is

not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.





References

1. Mosconi P, Radrezza S, Lettieri E, Santoro E (2019) Use of health apps and wearable devices: survey among italian associations for patient advocacy. *JMIR mHealth uHealth* 7(1):e10242
2. Islam SMR, Kwak D, Kabir MDH, Hossain M, Kwak KS (2015) The internet of things for health care: a comprehensive survey. *IEEE Access* 3:678–708
3. Evenson KR, Goto MM, Furlberg RD (2015) Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act* 12(1):159
4. Holland J (2016) Wearable technology and mobile innovations for next-generation education. IGI Global, USA
5. Peng C, Goswami P, Bai G (2020) A literature review of current technologies on health data integration for patient-centered health management. *Health Inform J* 26(3):1926–1951
6. Gyrard A, Patel P, Sheth AP, Serrano M (2016) Building the web of knowledge with smart IOT applications. *IEEE Intell Syst* 31(5):83
7. Jabbar S, Ullah F, Khalid S, Khan M, Han K (2017) Semantic interoperability in heterogeneous IoT infrastructure for healthcare. *Wirele Commun Mobile Comput* 2017:1–10
8. Laukkanen JA, Pukkala E, Rauramaa R, Mäkilä TH, Toriola AT, Kurl S (2010) Cardiorespiratory fitness, lifestyle factors and cancer risk and mortality in finnish men. *Eur J Cancer* 46(2):355–363
9. Kooiman TJM, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M (2015) Reliability and validity of ten consumer activity trackers. *BMC Sports Sci Med Rehabil* 7(1):24
10. Bunn JA, Jones C, Oliviera A, Webster MJ (2019) Assessment of step accuracy using the consumer technology association standard. *J Sports Sci* 37(3):244–248
11. Reda R., Carbonaro A. (2018) Design and development of a Linked Open Data-Based Web Portal for sharing IoT health and fitness datasets. In: Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good (Goodtechs '18), Bologna, Italy, November 28–30, 2018, pp. 43–48
12. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper, 2017
13. Reda R, Piccinini F, Carbonaro A (2018) Towards consistent data representation in the IoT healthcare landscape. In: Proceedings of the ACM DH'18: International Digital Health Conference 2018, Lyon, France, April 23–26, 2018, pp. 1–6
14. Reda R, Piccinini F, Carbonaro A (2018) Semantic modelling of smart healthcare data. *Adv Intell Syst Comput* 869:399–411
15. Carbonaro, A. (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
16. Alamri A (2018) Ontology middleware for integration of IoT healthcare information systems in EHR systems. *Computers* 7(4):51
17. Gay V, Leijdekkers P (2015) Bringing health and fitness data together for connected health care: mobile apps as enablers of interoperability. *J Med Internet Res* 17(11):e260
18. Kim HH, Lee SY, Baik SY, Kim JH (2015) MELLO: Medical lifelog ontology for data terms from self-tracking and lifelog devices. *Int J Med Inform* 84(12):1099–1110
19. Patel P, Gyrard A, Datta SK, Ali MI (2017) SWoTSuite: a toolkit for prototyping end-to-end semantic web of things applications. In: Proceedings of the 26th International Conference on World Wide Web Companion. International WWW Conferences Steering Committee 2017, Perth, Australia, April 03–07, 2017, pp. 263–267
20. Galkin M, Auer S, Kim HL, Scerri S (2016) Integration strategies for enterprise knowledge graphs. In: Proceedings of the 10th IEEE International Conference on Semantic Computing (ICSC 2016), Laguna Hills, CA, USA, February 4–6, 2016, pp. 242–245

21. Galkin M, Auer S, Vidal M, Scerri S (2017) Enterprise knowledge graphs: A semantic approach for knowledge management in the next generation of enterprise information systems. In: Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017), Porto, Portugal, April 26–29, 2017, pp. 88–98
22. Martino L, Ahuja S (2010) Privacy policies of personal health records: An evaluation of their effectiveness in protecting patient information. In: Proceedings of the 1st ACM International Health Informatics Symposium 2010 (IHI'10), Arlington, Virginia, USA, November 11–12, 2010, pp. 191–200
23. Roman D, Nikolov N, Putlier A, Sukhobok D, Elvesæter B, Berre A, Ye X, Dimitrov M, Simov A, Zarev M, Moynihan R, Roberts B, Berlocher I, Kim S, Lee T, Smith A, Heath T (2018) DataGraft: one-stop-shop for open data management. *Semant Web* 9(4):393–411
24. Belleau F, Nolin MA, Tourigny N, Rigault P, Morissette J (2008) Bio2RDF: towards a mashup to build bioinformatics knowledge systems. *J Biomed Inform* 41(5):706–716
25. Salvadores M, Alexander PR, Musen MA, Noy NF (2013) BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF. *Semant Web* 4(3):277–284
26. Walker RK, Hickey AM, Freedson PS (2016) Advantages and limitations of wearable activity trackers: considerations for patients and clinicians. *Clin J Oncol Nurs* 20(6):606–610
27. Sanchis-Gomar F, Lucia A, Yvert T, Ruiz-Casado A, Pareja-Galeano H, Santos-Lozano A, Fiuza-Luces C, Garatachea N, Lippi G, Boucharde C, Berger NA (2015) Physical inactivity and low fitness deserve more attention to alter cancer risk and prognosis. *Cancer Prev Res* 8(2):105–110
28. Nunez C, Clausen J, Jensen MT, Holtermann A, Gyntelberg F, Bauman A (2018) Main and interactive effects of physical activity, fitness and body mass in the prevention of cancer from the Copenhagen Male Study. *Sci Rep* 8:11780
29. Park W, Lee VJ, Ku B, Tanaka H (2014) Effect of walking speed and placement position interactions in determining the accuracy of various newer pedometers. *J Exerc Sci Fit* 12(1):31–37
30. Marshall MS (2012) Emerging practices for mapping and linking life sciences data using RDF—A case series. *J Web Semant* 14:2–13
31. Brunetti JM, Auer S, García R, Klímek J, Nečaský M. (2013) Formal linked data visualization model. In: Proceedings of International Conference on Information Integration and Web-based Applications & Services (IIWAS '13), Vienna, Austria, December, 2013, pp. 309–318
32. Dimou, A.; Vander Sande, M.; Slepicka, J.; Szekely, P.; Mannens, E.; Knoblock, C.; Van de Walle, R. Mapping hierarchical sources into RDF using the RML mapping language. *Semantic Computing (ICSC)*, 2014 IEEE International Conference on. IEEE, 2014, pp. 151–158
33. Riccucci, S., Carbonaro A., Casadei, G. (2005) An architecture for knowledge management in intelligent tutoring systems. In: Proceedings of IADIS International Congress on Cognition and Exploratory Learning in Digital Age, 2005, pp. 473–476
34. Carbonaro A. (2010). Towards an automatic forum summarization to support tutoring. In *International Conference on Technology Enhanced Learning* (pp. 141–147). Springer, Berlin, Heidelberg
35. Riccucci S., Carbonaro A., Casadei G (2007) Knowledge acquisition in intelligent tutoring system: A data mining approach. In *Mexican International Conference on Artificial Intelligence*. Springer, 1195–1205
36. Carbonaro A, Piccinini F, Reda R (2018) Integrating heterogeneous data of healthcare devices to enable domain data management. *J e-Learn Know Soc* 14(1):45–56
37. Carbonaro A, Ferrini R (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, pp. 270–288
38. Robinson LE, Holt TA, Rees K, Randeva HS, O'Hare JP (2013) Effects of exenatide and liraglutide on heart rate, blood pressure and body weight: systematic review and meta-analysis. *BMJ open* 3:e001986
39. McGuinness DL, Van Harmelen F (2004) OWL web ontology language overview. *W3C recommendation* 10(10):2004

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Roberto Reda¹  · **Filippo Piccinini²**  · **Giovanni Martinelli²**  ·
Antonella Carbonaro¹ 

Roberto Reda
roberto.reda@unibo.it

Filippo Piccinini
filippo.piccinini@irst.emr.it

Giovanni Martinelli
giovanni.martinelli@irst.emr.it

¹ Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

² IRCCS Istituto Romagnolo Per Lo Studio Dei Tumori (IRST) Dino Amadori, Meldola, FC, Italy