Advancing Quantified-Self Applications Utilizing Visual Data Analytics and the Internet of Things

Dimitris Chousiadas, Andreas Menychtas, Panayiotis Tsanakas, Ilias Maglogiannis

Advancing quantified-self applications utilizing visual data analytics and the Internet of Things

Dimitris Chousiadas[1], Andreas Menychtas[2], Panayiotis Tsanakas[3] and Ilias Maglogiannis[2]

[1] National Technical University of Athens, Athens 15773, Greece
dim.chousiadas@gmail.com
[2] University of Piraeus Department of Digital Systems, Piraeus 18532, Greece
{amenychtas, imaglo}@unipi.gr
[3] Greek Research and Technology Network (GRNET), Athens 11523, Greece
tsanakas@grnet.gr, panag@cs.ntua.gr

Abstract. The exponential growth of the number and variety of IoT devices and applications for personal use, as well as the improvement of their quality and performance, facilitates the realization of intelligent eHealth concepts. Nowadays, it is easier than ever for individuals to monitor themselves, quantify and log their everyday activities in order to gain insights about their body performance and receive recommendations and incentives to improve it. Of course, in order for such systems to live up to the promise, given the treasure trove of data that is collected, machine learning techniques need to be integrated in the processing and analysis of the data. This systematic and automated quantification, logging and analysis of personal data, using IoT and AI technologies, has given birth to the phenomenon of Quantified-Self. This work proposes a prototype decentralized Quantified-Self application, built on top of a dedicated IoT gateway, that aggregates and analyses data from multiple sources, such as biosignal sensors and wearables, and performs analytics and visualization on it.

Keywords: Quantified-Self, Internet of Things, Visual Data Analytics, Personal Informatics, Sensors, Activity Tracking, eHealth, Containers.

1 Introduction

The technological developments of the last decade contributed to an abundance of commercially available and affordable biomedical sensors, activity trackers and wearables. Moreover, these sensors come with specially developed smartphone applications, which enable the users to collect data about themselves, in a systematic and automated manner, and have access to reports and analyses, providing insights and valuable feedback.

The types of data that is collected in the context of Quantified-Self are virtually countless, including but not limited to: health, physical activity, nutrition, psychological and mental state or environment [26]. For each of these categories, several applications exist that focus on the collection and analysis of relevant data, all of them aiming at highlighting the individual characteristics of the user's lifestyle, identifying his or her strong and weak points and triggering desirable behavior change. Machine learning and big data are technologies that are applied in this area to facilitate that analysis of the data and the creation of personalized knowledge for users and medical experts [8].

According to research, the mere monitoring and logging of values, such as body weight, is effective in itself, by making the user conscious of its importance and in turn, leading to lifestyle changes [17]. The tracking of values over time, additionally, produces insights regarding the user's progress, highlighting the direct relation between their efforts and the results (e.g. systematically going to the gym leads to a decrease of body fat percentage [20]). State-of-the-art solutions include mobile applications for GPS tracking of cardio activities, websites for tracking body weight, fat percentage and bone mass using connected body scales, mobile apps for blood pressure monitoring, apps with integrated activity tracking [24].

Research has shown, however, that over 50 percent of people using similar applications have given up on them after a few months [11]. Some of the reasons include absence of a holistic approach to data collection and analysis, and lack of customization capabilities [19]. The constant emergence of new applications and corresponding wearable devices, usually incompatible with the previous ones, hardens the adoption of a vendor-free solution like the one proposed in this work. Furthermore, there exist privacy concerns in current Quantified-Self applications, since, typically, the collected data is permanently stored in the cloud, giving ownership and, thus, control of it to the application provider. The greater the sensitivity of the data, the greater the risk for the user, in case of data loss, data manipulation or hacker attacks. Moreover, a lot of providers, in order to maintain the free-of-charge status of their Quantified-Self services, sell the data to third parties (e.g. for advertising or marketing purposes) [13].

The proposed Quantified-Self approach aims to provide the technical and functional background to support data collection from multiple sources, in the context of health and activity tracking, the analysis of that data and the visualization of the results, while taking into consideration the aforementioned shortcomings of the state-of-the-art applications. On the one hand, it allows the user to combine on the same platform several devices, resolving the compatibility issues by exploiting the advancements of Bluetooth communication technologies, and particularly of BLE [15]. On the other hand, being built on top of the innovative software stack of AGILE [1] and deployed on a Raspberry Pi, instead of the typical hybrid deployment in smartphone and cloud, it gives ownership of the data over to the user.

The rest of the paper is structured as follows: Section 2 describes the related work in the area of Quantified-Self while section 3 presents the proposed solution and the implementation of the system. The resulting application, with the data analysis and visualization features is presented in section 4. Finally, section 5 discusses the advancements of this approach compared to other approaches, concludes this work and highlights the future plans.

2 Related Work

A Quantified-Self application, such as the one presented here, aims at, not simply quantifying and gathering data, but also exploiting data analysis methodologies and tools to give feedback to the user and promote behavior change. F. Bentley et al. [3] have studied the ways such change can occur, particularly in the context of health and wellbeing, highlighting the importance of the general (environmental and personal) context of the user's life, a conclusion that has also been reached by Choe et al. [6], who have shown that not tracking context and triggers leads to decreased chances to improve outcomes. This necessitates a holistic approach to data collection, in order to capture and quantify the whole spectrum of this context while in parallel, given the multifaceted nature of the user's everyday life, sophisticated data mining techniques need to be used, in order to uncover hidden, meaningful patterns in it. F. Bentley's et al. research, moreover, explores how users perceive and react to recommendations about behavior change, suggesting that the use of natural language, instead of charts or numbers, can sometimes be more effective.

With regards to the ways IoT data can be harvested, Chen, Feng, et al [5] have presented an overview of the various data mining techniques, applied in the context of IoT. What distinguishes IoT systems, concerning data mining, is the requirement for real-time data processing, as well as processing at the edge of the network (Edge Analytics). As an example, such requirements exist in telemonitoring systems, such as the ones presented by C. Panagopoulos et al. [23] and K. Aguilar et al. [2], where doctors need to monitor patients and receive notifications in real time. Strongly connected with the above is the research of J. Li et al. [14], which investigates the potential IoT applications in a more general context than that of health and medical care. More concretely, through the lens of fog computing the researchers highlight the ways Smart Living can be advanced, in various aspects of everyday life (work, entertainment, energy consumption and, naturally, health) through the quantification and analysis of data.

Regarding the specific sensors that are used, as well as the architecture of the sensor network and the flow of collected data, the works of A. Kor et al. [12] and A.H.T.E. De Silva et al. [7] make obvious the relevance and applicability of the Internet of Things and Quantified-Self concepts in healthcare and mHealth. Finally, B. D. Weinberg et al. [27] have examined the privacy concerns that accompany the transition from Web 2.0 to Internet of Things. Conflicts of interests between users and organizations have elevated data ownership to a crucial issue. In the context of IoT, where data is generated automatically and continually, the question of who owns it is, indeed, of paramount importance.

3 System Architecture

The Quantified-Self concept is targeting data acquisition on aspects of a person's daily life in terms of inputs (e.g. food consumed, quality of surrounding air), states (e.g. mood, arousal, blood oxygen levels), and performance (mental and physical activities) through a modern, health centric, social and mobile enabled, communication platform that resides in the gateway (in terms of collecting and visualizing data).

However, the ecosystem of biosignal sensors and activity trackers is very diverse from both business and technical perspectives. The vendors have also different business and market strategies, and in many cases these strategies are reflected in the devices themselves in technical and operational level. The open source designs for such devices that can be used for producing or communicating with them are uncommon. The hardware and firmware of the majority of the devices is proprietary and closed source, hence establishing communication with the device as a third-party required guidance from the vendors either by providing access to the communication protocols or through libraries / APIs. Therefore, what is usually happening today, is that the users are obliged to use each wearable and sensor with different, companion applications, which store the date to independent systems.

The unique characteristic of the proposed approach is that the Quantified-Self application is deployed on top of an IoT Gateway [18], which eliminates the need for additional applications or hardware. Wearable activity trackers and medical sensors automatically communicate with the gateway whenever within range, and offload the most recent data. However, for the wearables that the manufacturers provide only API libraries for Android and iOS, or specific applications for synchronizing the data with manufacturers' cloud platforms, neither of which is compatible with the proposed setup. In these cases, the activity data offloading is achieved via Internet and the public APIs of the cloud platform.

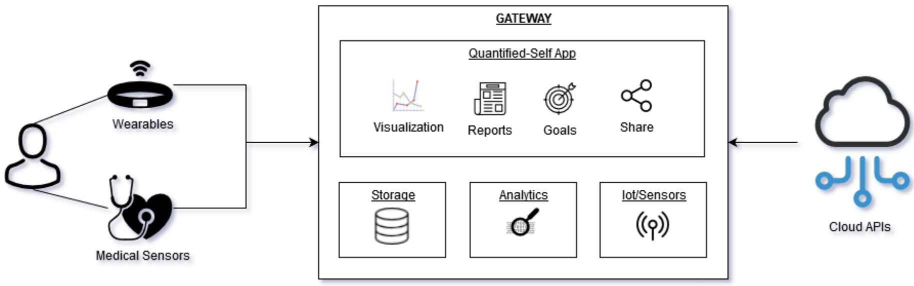

Fig. 1. Quantified-Self Concept

Fig. 1 presents the general architecture for the proposed Quantified-Self application. Each user connects the various activity tracking devices and biosignals sensors (such as oximeters, blood pressure monitors or glucometers), to monitor their daily physical activity and physical condition. The user is able to visualize and manage the locally stored data and create reports. In parallel, the data are processed and analyzed on the gateway, with personalized recommendations being sent to the users, encouraging them to reach their physical activity goals. In order to address the requirements for modularity and extensibility, the architectural model of the application is based on containers. More specifically, four main containers are deployed on the gateway to provide the

required functionality of the application: a) *IoT/Sensors* for managing the communication with the various sensors and acquiring the biosignal measurements, b) *Storage* for storing the data locally, c) *Analytics* for the periodic and ad-hoc analysis of the data in order to create valuable knowledge for the users' health condition and activity and finally d) *the Quantified-Self Web Application* which coordinates the previous containers and provides a UI with modern reporting and data visualization functionality.
All components of the application (the core application, database and Shimmer) run on a Raspberry Pi, on top of the AGILE Software Stack [18], on a Linux based operating system. The container-based deployment approach of AGILE was adopted, ensuring the virtual isolation of the components from each other and providing the required flexibility for the maintenance and extension of the system with additional features.

3.1 IoT Sensors Integration

The AGILE IoT Gateway software stack is a flexible architecture built for single board computers running independently or on the edge of the network. By providing developers ready, short distance networking abilities, such as Bluetooth or Zigbee, it allows acquisition from local sensors, view and local storing of data. For the integration of the Quantified-Self biosignal sensors and activity trackers, the respective components of AGILE were used. Using the harmonized REST API, the integration of any sensor is simplified in the cases where specific commands or operational workflows are required, custom device drivers are implemented. This approach follows a registration / publish / subscribe device model. Each device needs to be registered before use and the application can subscribe to the sensor data endpoints in order for published data to be streamed to the user. This model works well with BLE devices, which use GATT notifications [4] to stream data to the host.

3.2 Data Management

Attempting to combine on the same platform data from different medical devices, such as oximeters, glucometers, smart scales and blood pressure monitors, introduces significant complexities regarding the storage and manipulation of that data. Integrating, furthermore, data from the cloud APIs of Fitbit [9] and GoogleFit [10], only increases these complexities. The homogeneous and consistent manipulation of the collected data, however, is a necessary prerequisite for any meaningful processing to take place in the system. To this end, the tools developed by the open-source project Open mHealth [22] proved particularly useful. These tools include, a collection of schemata that define the structure of health data. Specifically, these schemata standardize the representation of health data following the syntactic rules of JSON, i.e. key-value pairs, defining the keys to be used for all types of health data. For the purposes of the proposed Quantified-Self application this standardization facilitates the easy and smooth integration of the data, regardless of the source, and simplifies the architectural choices for the database and the REST API. In the case of the medical sensors that communicate with the gateway through BLE, the conversion of the data to JSON objects that comply with the Open mHealth rules takes place during that communication.

3.3 Data Analytics

The amount of the activity and medical data that is collected gives ample opportunity for the analysis of that data and the discovery of patterns in it, providing valuable feedback to the user. The analytics component is based on the python tools of Scikit-learn [25], Matplotlib [16] and Numpy [21] which have been deployed on a specific container configured to communicate with the database and the web application. Scikit-learn, specifically, is a well-known machine-learning library for Python, featuring classification, regression and clustering algorithms, including the k-means algorithm that was extensively used in the proposed application. Numpy is numerical library for Python that provides, primarily, support for multi-dimensional arrays and matrices, designed to interoperate with Scikit-learn. The data visualization was realized utilizing Matplotlib, which is a plotting library for Python, along with the drawing capabilities of the JavaScript programming language.

Since all data are stored locally, the data analytics component periodically, or upon user's request, processes the activity and biosignal data in order to create reports and aggregated information to be presented in the web app. In the current implementation data clustering and correlation techniques have been used, and the results are also exploited to provide motivational messages to the users. Regarding the clustering techniques, data cleaning and pre-processing was necessary. This included removing the outliers from the datasets, normalizing the measurements values, as well as adding potential missing values, either by replacing them with the mean value of the dataset or by filling in the respective values from the nearest data vector in the Euclidean space. Once the data cleaning was done, the data was aggregated, in most cases by day, and the k-means algorithm was applied on that aggregated data.

3.4 Quantified-Self Web Application

The user interacts with the system through a web application, which is also hosted on the IoT Gateway and is only accessible through the local home network and does not expose itself or the user's data to public networks or the Internet. Through the application, the users are able to manage the various IoT sensors (e.g. register and initialize them) and the stored data by communicating with the respective components of the system. The web application also visualizes the user's activity data, collected both from local sensors (if available) and the APIs of GoogleFit and/or Fitbit. The activity consists of the number of steps taken each day, the duration of physical activity every day, as well as the calories burned. Examples of such visualizations are shown in **Fig. 2** and **Fig. 3**, where the user can see the number the steps taken during the last 7 days or the average duration of the user's physical activity for each day of week for the last year.

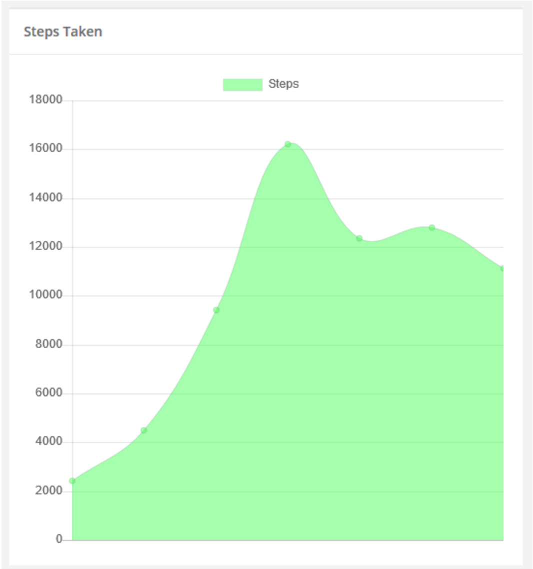

Fig. 2. Number of steps taken the last 7 days

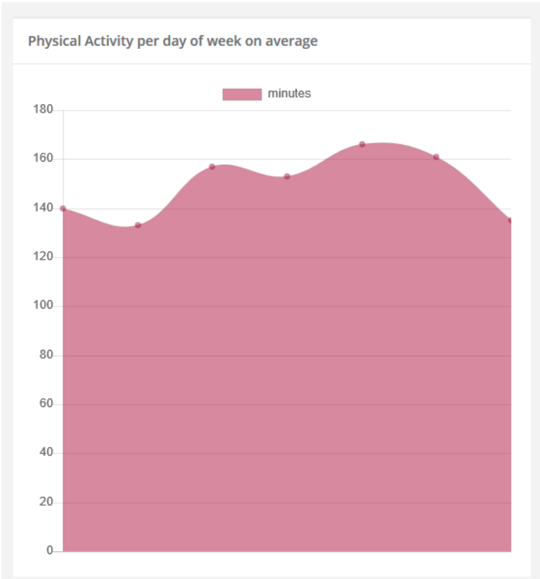

Fig. 3. Duration of physical activity on average per day of week

Another category of data is the biosignals, namely heart rate, blood pressure, oxygen saturation, glucose levels and body weight. These are collected through specific sensors that communicate with the AGILE Gateway or directly, as user input. Similar visualization exists for that second category as well.

On the home page of the application, the user can find general results of the statistical analysis of the data, such as the ones shown in **Fig. 4** and **Fig. 5**. Based on these, relevant messages appear to encourage and motivate the user. Furthermore, the system informs the user of how the overall community of users is doing, and he or she compares to it. Finally, there is also the option of setting goals, regarding the performance of the user. Ultimately, these features aim to maximize the user's commitment and bring about improved performance.

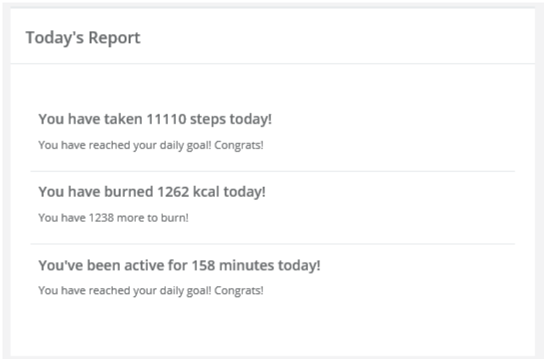

Fig. 4. Information about today's activity presented in the Homepage

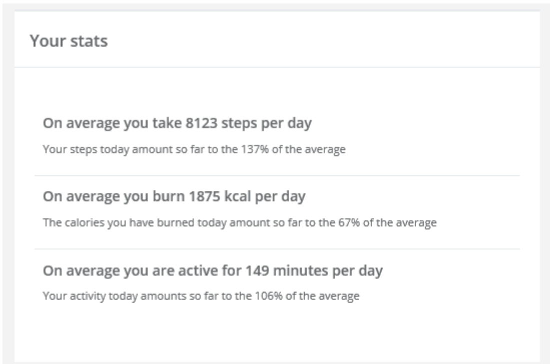

Fig. 5. Information about the average activity of the user, also presented in the Homepage

4 Experimentation

The proposed Quantified-Self application was provided to real users in order to assess the visual data analytics functionality. The biosignals data were analyzed using well-known k-means algorithm. As an example of this, in particular, the heartbeat and oxygen saturation data have been clustered in three clusters. Specifically, each data point represents the average heartbeat and oxygen saturation measurements in a specific day and the choice of three clusters corresponds to the intuition that, for any given user, there are days with good measurements (green cluster), days with medium measurements (red cluster) or days with bad measurements (blue measurements). Furthermore, using the elbow method, the choice of three clusters was shown to be close to optimal, and it was, thus, selected due to its to intuitive interpretation. Once the clustering has been made, each day can be assigned to one cluster, informing, thus, the user, of the quality of his/her measurements that day. The results of the clustering can be presented to the user with a pie chart, as in **Fig. 6** with the percentage of the user's measurements in each cluster.

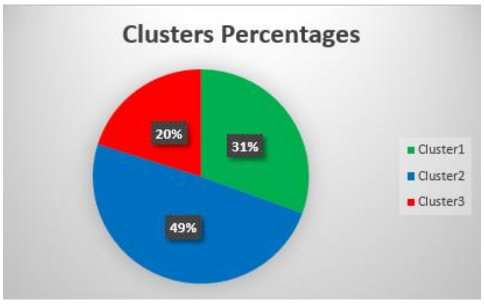

Fig. 6. Percentage of measurements in the 3 clusters

Another example of how clustering can provide useful insights to the user is presented in **Fig. 7**. Specifically, each day of week is broken down in the percentages of the measurements of that day that belong to each one of the 3 clusters.

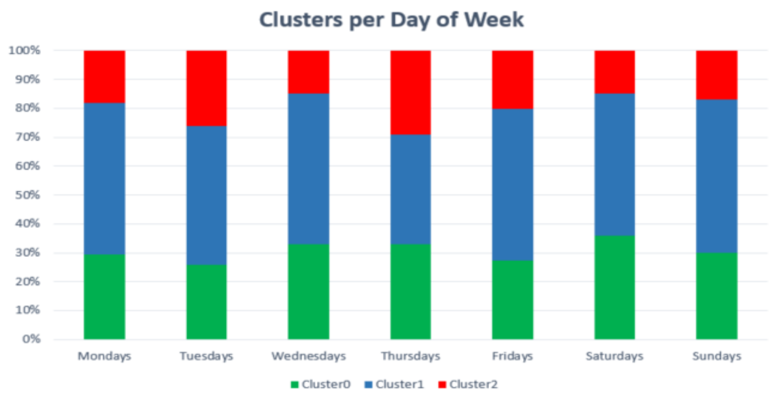

Fig. 7. Clusters per day of week

Moreover, a visualization is also included in the web application in order to provide valuable information to the user presenting the results of the clustering analysis as a time series. Concretely, to each one of the, say, last 30 days, a cluster 1 to 3 is assigned, with cluster 1 representing good measurements, cluster 2 medium measurements and cluster 3 bad measurements. It is possible to see, then, how the clustering changes or, in other words, how the quality of the user's measurements changes between days. In **Fig. 8** the visualization of such an analysis can be seen. Specifically, the blue columns represent the cluster values for each day in the last month, whereas the orange ones present the delta of these values, namely, how small or big the change is between days. In this way, significant changes, i.e. anomalies can be traced.

Along with the clustering, the application computes the correlation between the two datasets, using the Pearson Correlation Coefficient, in order to indicate the strength of the linear relationship between the data (for the clustering example presented above, the coefficient is equal to -0.22). Generally, the correlation coefficient can be calculated for any combination of the data types in the application, enabling the user to receive feedback in the form of natural language phrases, such as: “On days when you run more than 20 minutes, your oxygen saturation level increases”.

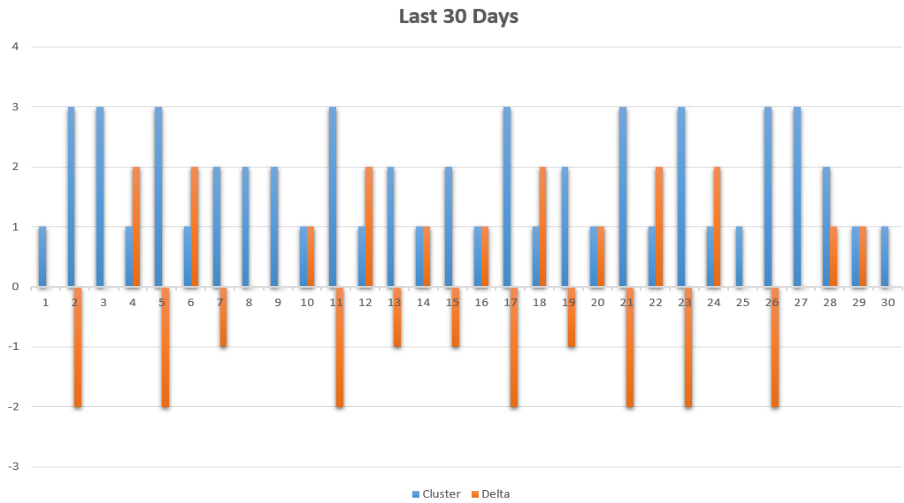

Fig. 8. Time series of the clusters of last 30 days: the blue bars represent the cluster to which each one of the last 30 days belongs, whereas the orange bars represent the change of clusters between days

5 Discussion and Conclusions

The competitive advantages of the proposed Quantified-Self application are twofold. The main innovation, on the business level, is the fact that the data remain on the user’s gateway, while data processing and analytics are provided locally. On the technical level, the infrastructure of the AGILE Gateway supports communication with and integration of several devices on the same platform. These two, combined, bring several benefits for the end-user, as well as for the developers’ community. Furthermore, the visualization that the application employs provide valuable feedback to the user. The combination of different charts, shapes, coloring and phrasing of the results aim at enabling the users to intuitively understand and interpret their behavior and performance.

The application proposes a fully automated solution that requires minor engagement of the user, making it suitable for people unfamiliar with technology. What further makes the solution user-friendly and advances its ease-of-use is the fact that the underlying IoT gateway can be built using affordable, commodity hardware. Moreover, the native support for modularity, extensibility and high customization, without the need to speed effort across the different layers and components of the gateway, is one of the most important benefits for the Quantified-Self application allowing for its continuous evolution and adaptation following the users' growing requirements and the new technological solutions and trends in the domain.

The application will be further extended by employing more sophisticated gamification techniques to advance the goal setting and data sharing features, as well as additional machine learning methods, like classification or regression, which would be used to predict value for the user, such as future body weight. Furthermore, there is also room for the integration of more interactive and informative visualization tools, which would build on top of the existing ones.

Acknowledgements. Dr. A. Menychtas acknowledges the Greek State Scholarship Foundation (IKY). This research was implemented with a Scholarship from IKY and was funded from the action "Reinforcement of postdoctoral researchers" of the programme "Development of Human Resources, Education and Lifelong Learning", with priority axes 6,8,9 and it was co-financed by the European Social Fund-ESF and the Greek State.

References

1. AGILE-IOT, H2020 EU Research Project: http://agile-iot.eu
2. Aguilar, K.M., Campbell, R.S., Fiester, A., Simpson, R.L. and Hertel, C.: Bringing care home: how telemonitoring can expand population health management beyond the hospital. Nursing administration quarterly, 38(2), pp.166-172 (2014)
3. Bentley, F., Tollmar, K., Stephenson, P., Levy, L., Jones, B., Robertson, S., Price, E., Catrambone, R. and Wilson, J.: Health Mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. ACM Transactions on Computer-Human Interaction (TOCHI), 20(5), p.30 (2013)
4. Bluetooth GATT Specifications: https://www.bluetooth.com/specifications/gatt
5. Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A.V. and Rong, X.: Data mining for the internet of things: literature review and challenges. International Journal of Distributed Sensor Networks, 11(8), p.431047 (2015)
6. Choe, E.K., Lee, N.B., Lee, B., Pratt, W. and Kientz, J.A.: Understanding quantified-selfers' practices in collecting and exploring personal data. In: Proceedings of the 32nd annual ACM conference on Human factors in computing systems (pp. 1143-1152). ACM (2014)
7. De Silva, A.H.T.E., Sampath, W.H.P., Sameera, N.H.L., Amarasinghe, Y.W.R. and Mitani, A.: October. Development of a wearable tele-monitoring system with IoT for bio-medical applications. In: Consumer Electronics, 2016 IEEE 5th Global Conference on (pp. 1-2). IEEE (2016)
8. Fawcett, T.: Mining the quantified self: personal knowledge discovery as a challenge for data science. Big Data, 3(4), pp.249-266 (2015)
9. Fitbit Official Website: https://www.fitbit.com/eu/home
10. GoogleFit Webiste: https://www.google.com/fit/

11. Ledger, D. and McCaffrey, D.: Endeavour Partners Report: Inside Wearables: How the Science of Human Behavior Change Offers the Secret to Long-term Engagement. 2016
12. Kor, A.L., Yanovsky, M., Pattinson, C. and Kharchenko, V.: SMART-ITEM: ioT-enabled smart living. In: Future Technologies Conference (FTC) (pp. 739-749). IEEE (2016)
13. Leibenger, D., Möllers, F., Petrlic, A., Petrlic, R. and Sorge, C.: Privacy challenges in the quantified self movement–an EU perspective. Proceedings on Privacy Enhancing Technologies, pp.315-334 (2016)
14. Li, J., Jin, J., Yuan, D., Palaniswami, M. and Moessner, K.: EHOPES: Data-centered Fog platform for smart living. In: Telecommunication Networks and Applications Conference (ITNAC), 2015 International (pp. 308-313). IEEE (2015)
15. Mackensen, E., Lai, M. and Wendt, T.M.: Bluetooth Low Energy (BLE) based wireless sensors. In: Sensors, 2012 IEEE (pp. 1-4). IEEE (2012)
16. Matplotlib, plotting library for the Python: https://matplotlib.org/
17. McGrath, M.J. and Scanaill, C.N.: Wellness, fitness, and lifestyle sensing applications. In: Sensor Technologies (pp. 217-248). Apress, Berkeley, CA (2013)
18. Menychtas, A., Doukas, C., Tsanakas, P. and Maglogiannis, I.: A Versatile Architecture for Building IoT Quantified-Self Applications. In: IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS) pp. 500-505. IEEE (2017)
19. Menychtas, A., Tsanakas, P. and Maglogiannis, I.: Automated integration of wireless biosignal collection devices for patient-centred decision-making in point-of-care systems. Healthcare technology letters, 3(1), p.34 (2016)
20. Munson, S.A. and Consolvo, S.: Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In: Pervasive computing technologies for healthcare (PervasiveHealth), 6th international conference on (pp. 25-32). IEEE (2012)
21. Numpy, package for scientific computing with Python: http://www.numpy.org/
22. Open mHealth: http://www.openmhealth.org/
23. Panagopoulos, C., Malli, F., Menychtas, A., Smyrli, E.P., Georgountzou, A., Daniil, Z., Gourgoulianis, K.I., Tsanakas, P. and Maglogiannis, I.: Utilizing a Homecare Platform for Remote Monitoring of Patients with Idiopathic Pulmonary Fibrosis. In: GeNeDis 2016 (pp. 177-187). Springer, Cham (2016)
24. Pcmag.com: The Best Fitness Trackers of 2017, Available online at: http://www.pcmag.com/article2/0,2817,2404445,00.asp
25. Scikit-learn, machine learning in Python: http://scikit-learn.org/stable/
26. Swan, M.: The quantified self: Fundamental disruption in big data science and biological discovery. Big Data, 1(2), pp.85-99 (2013)
27. Weinberg, B.D., Milne, G.R., Andonova, Y.G. and Hajjat, F.M.: Internet of Things: Convenience vs. privacy and secrecy. Business Horizons, 58(6), pp.615-624 (2015)