Greenolive: an Open Platform for Wellness Management Ecosystem
Liangzhao Zeng, Pei-Yun Hsueh, Henry Chang
{lzeng, pshsueh, hychang} @us.ibm.com
IBM T.J. Watson Research Center, U.S.A

Abstract— In this paper, we present Greenolive, an open platform for wellness management ecosystem. Wellness management applications, which facilitate preventive care and chronic disease treatments, are considered as a key component to enhance healthcare quality and reduce healthcare cost. Currently, most of the wellness management applications are device-oriented, stand-alone software and lack of open APIs that allow new value-added applications to be developed rapidly. Further, these wellness applications have not fully utilized the collected wellness monitoring data to generate new knowledge that can help people to further improve their wellness status. In this paper, we advocate an open platform that provides services that are essential to wellness management, and manages the services to be developed. Further, we adopt an elastic infrastructure to deal with scalability issues. We expect independent software vendors to develop new wellness management applications using the provided services and to deploy them on the platform. We believe such an ecosystem can greatly promote wellness management applications, much like we witness in other application domains, e.g., social network, enterprise resource management.

I. INTRODUCTION
Wellness management, which aims to manage people’s lifestyle, is key for preventive care and chronic disease treatments that help people maintain and improve their health. For example, the cornerstone of Type II diabetes treatment consists of lifestyle modification and exercise and weight management[3], which can be facilitated and enhanced by wellness management applications. With the recent advances in personal wellness monitoring devices and living sensors, wellness management applications are now well-positioned to provide awareness on wellness status and treatment progress, generate alerts of potential risk, and make suggestions on how to maintain and improve healthy lifestyle. Specifically, these applications are developed based on collected information about vital signs (e.g., heart rate, blood pressure), nutrition, physical exercise, and living environments. It is considered as an efficient and low cost solution for improving healthcare quality. However, current wellness management services/applications are not as popular as they should be, as there are some key weaknesses. First, most of the current solutions are device-oriented and PC-based one-fits-all applications. As wellness management applications should be highly personalized, the one-fits-all paradigm usually cannot satisfy the unique requirements of different users. For example, glucose monitoring should be different for Type I and Type II diabetes. Also, these applications are tightly integrated with the provided devices.

Such an application per device paradigm has imposed burdens on users to coordinate the many disintegrated/disconnected applications. In an ideal case, these applications can work collaboratively to provide comprehensive services for end users. Further, these PC-based applications lack of open API that can enable the development of more value added services. In contrast, in other areas such as social network[2] and customer relationship management[3], the platform operators do provide easy accessibility (provisioned as web applications) and open APIs, and independent software vendors (ISVs) have developed many new popular services and business models.

In fact, the progress in these application areas (e.g., social network) gives us directions on how wellness management applications should be evolved. On the one hand, it is important to provide an open platform to build ecosystems. On the other hand, developing a platform for wellness management has its own challenges:

- **Wellness data collection.** There are various data sources from different kinds of devices/sensors for wellness management applications, which include device/personal identification, vital signs, nutrition, physical activity, living environments, etc. Even for the same type of devices, different vendors may adopt different data formats to represent the measurement. Therefore, a unified ontology framework is required so that data in different formats can be transformed in a unified format and semantics. Another issue is about data collection. Data from various devices/sensors or media (e.g., data feed from Web) are collected with different collection policies. For example, when collecting data from vital signs monitoring devices, i.e., blood pressure, some clinical guidelines need to be followed, in order to guarantee the data quality. It is important for the platform to facilitate the data collection process (e.g., data format/semantics transformation), so that data can be used by different applications/services.

- **Realtime wellness monitoring.** Providing wellness awareness for users is a key issue, as it can give end users incentives to continue using the wellness management applications. For example, when wellness measure data are uploaded, certain feedbacks about the user’s wellness status should be provided. Technically, it is an event processing problem, with each data collection session considered as an event occurrence. Therefore, the platform needs to provide a programming model that allows ISVs to specify and execute the event processing logic. Once the event processing logic is deployed, how to efficiently execute them is a challenge,
given the fact that overall volume of events can be very high with the number of users increasing. Further, the event processing logics are highly dynamic, as the monitored entity (human)’s activity and status are very dynamic. Therefore, dynamicity support is a key issue for successful wellness monitoring.

- **Analytics for wellness compliance and progress.** With the collection of wellness data and results from realtime event processing, the platform can accumulate a very rich data set for analytics to generate wellness evidences. The generated evidences can be either used in realtime monitoring as new event processing logic, or provide as suggestions for end users to change their lifestyles for better wellness. In this aspect, the platform needs to provide tools that allow ISVs to specify and execute Extraction-Transform-Load (ETL) logic and to define evidence mining tasks. Also, given the fact that data is stored in distributed locations, a scale-out solution may be required. Furthermore, the platform also needs to provide mechanisms to use evidences generated. It should be noted that when the end-users’ data are used for analytics, appropriate access control is required, for protecting the privacy of end users.

In our project, to tackle these challenges we develop an open platform, dubbed Greenolive, for wellness management ecosystem. The distinct features of our platform are:

- **Open platform for ecosystem as a service.** We adopt the software-as-a-service paradigm that enables ISVs to develop wellness management applications. Such an approach requires the platform to provide some services (APIs) that are essential to develop wellness management applications. To enable an open platform, we adopt Web services (i.e., restful API) to provision the wellness management functionalities.

Here, we provide four categories of services that are required to support most of the wellness management scenarios, which include (i) data transformation & routing service that facilitates information flow on the platform; (ii) wellness monitor service that processes information in realtime fashion; (iii) wellness analytics that provides deeper understanding on people’s wellness; (iv) personal wellness record & knowledge repository that persists wellness information and provides information access APIs.

- **Elastic cloud architecture for scalable service provision and execution.** By supporting ISVs to develop new wellness management applications on this platform, we expect very dynamic workload. Also, the different services come with different workload requirements. For example, the workload of data transformation & routing services is depended on the number of users and associated devices and the frequency of data collection. While for wellness monitor services, workload is mainly depended on event volume and complexity of the event processing logic. These two services consume resources of network bandwidth, memory and CPU and impose realtime processing requirement. The wellness analytics services consume additional I/O and storage and do not always work in realtime. In our work, we adopt an elastic cloud architecture to address the scalability and dynamic workload issues.

In this paper, we will present an overview of the platform and mainly focuses on the monitor and analytics features. Due to the limitation of space, the detailed discussion of service management is out of scope of this paper. The rest of paper is organized as follows. Section 2 provides the overview of our platform. Section 3 describes the wellness monitoring framework. Section 4 illustrates the wellness analytics services. Section 5 reviews related work and Section 6 concludes the paper.

## II. GREENOLIVE PLATFORM

Fig. 1 illustrates the simplified Greenolive platform architecture. There are four main components, namely (i) Data Transformation & Routing Service; (ii) Wellness Monitor Service; (iii) Wellness Analytic Service; and (iv) Wellness Record & Knowledge Repository. With these components that build on top of a cloud infrastructure, developers can readily create two kinds of portals essential to personal wellness, namely (i) Wellness Management Portal that connects with the devices/sensors and provides end users wellness services based on the collected data, and (ii) Wellness Care Portal for care assistants. It should be noted that Wellness Management Portals can be deployed on either PCs or Smartphones. In the following subsections, we will discuss information flow, present wellness application development cycle, and explain how wellness management is realized, using a scenario.

### A. Information Flow

Our platform supports two kinds of information flows, namely event flow and data flow. In our platform, sensors/devices are connected to the information gateway that Wellness Management Portals run on. When end users are using wellness services, data are collected from sensors/devices following pre-specified data collection policies. Data collection policies are formulated as Event Condition Action (ECA) rules in our system. An event can be (i) Time-based trigger, e.g., every day at 10 o’clock, every 5 minutes for a week. Such events usually are used to collect vital signs data; (ii) Life activity event, e.g., before/after every meal. Such events usually are used to initiate vital signs collection that is driven by clinical guidelines. For example, to trace the wellness status of a diabetic patient, it is import to record glucose level before and after meal; and (iii) Event triggered by measurement result of vital signs. Such events are used to trigger follow-up measures when abnormal vital signs are detected.
For example, blood pressure measurements indicate hypertension, ECG measurements may be required. Combined with user identifiers, all collected data is considered as raw data (which may be encrypted for privacy/security concerns) that would be sent to the Data Transformation & Routing Service. When the service receives raw data, it transforms the information into a standard format with semantics annotation. In our platform, we adopt Continua [1] as the target transformation format to be conformed to. Further, the service determinates whether an event should be generated and routed to the Wellness Monitor Service for further processing. After the monitor service processes the event, a realtime feedback can be generated for related end users. Further, wellness event can be used by the analytic service for prediction scoring. Such an information flow is considered as the event flow. Meanwhile, the routing service forwards the wellness data in the Wellness Record & Knowledge Repository for later access. Integrated with other data sources such as previous personal health records, the accumulated wellness data can then lend support to wellness evidence mining as parts of the Wellness Analytics Service. Eventually, the new wellness knowledge will be delivered to end users. This is considered as the data flow of the system.

B. Service Development Cycle

Greenolive provides open APIs that allow ISVs to develop new wellness management applications. A typical wellness application consists of data transformation & routing services, wellness monitor services and wellness analytic services. Therefore, developing new applications includes the following three aspects of efforts.

1) Developing Data Transformation & Routing Service

The first step to develop a new wellness management service is identifying related data sources. If new devices/sensors are required, the developer can customize the data transformation services to transform the raw data from standardized format. Further, the developer needs to model the data collection policies and deploy them as part of the Wellness Management Applications. The policy specifies when and how the data should be collected. For example, blood glucose levels should be measured before and after every meal. It should be noted that developers also need to decide that whether collected raw data need to be processed in realtime fashion or not. If realtime monitoring is necessary, event instances are generated and routed to Wellness Monitoring Services.

2) Programming Monitoring Logic

The second step to develop a new service is programming monitoring logic. The platform provides APIs that allows developers to (i) subscribe to the related events; (ii) filter the un-interested events; (iii) correlate event to associated monitoring entities. In our platform, notion monitor context is used to represent the monitor entities, i.e., the wellness status of end users. A monitor context instance is identified by a unique user ID and persisted in personal wellness record; (iv) update the monitored status of entities, and (v) generate and delivery alerts. It should be noted on our platform, the event processing logic is highly dynamic, owing to the adjustment of the treatment plan and wellness goal and this dynamicity of wellness services.

3) Developing Analytic Service

The third step to develop the wellness management service is creating analytics applications. The platform provides an API for healthcare applications to (i) integrate heterogeneous data source (Sense), (ii) draw predictions by applying or extending models in a repository (Predict), and (iii) trigger proper responses (Respond). The ultimate goal is to enable any ISV to use the API and the Sense-Predict-Respond framework to implement their services and exchange information with 3rd party applications.

In the Sense stage of (i), a developer first creates a template that records steps needed for generating features from selected data sources. In particular, the template calls functions from API to perform three main tasks: privacy control, data cleaning, and feature extraction. Privacy control functions verify whether a user can access all the requested data fields by invoking policies which determine who have the authority to access what part of the data and what combination of the data fields should never be accessed.

---

![Fig 1. Simplified Greenolive Platform Architecture](image-url)
together. Data cleaning functions maintain rules of handling missing values, and data that are uncertain and inconsistent. Feature extraction functions specify how to recognize semantic entities and their syntactic links from unstructured information (such as clinical notes) and how to transform the extracted information into standardized features.

In the Predict stage of (ii), a developer locates a model repository, identifies a set of models needed for meeting the requirement of the analytic application, chooses one model from the repository, and performs analysis based on a target patient’s feature set. If none of the existing models meet the requirement, the developer applies the Create function to train a new model, based on the modeling method in use and the feature set in a modeling schema. If the selected model needs to be adapted with new data, such as the recent history of a target patient, the developer applies the Update function to re-train the model based on the specified method and the feature set in the modeling schema. To enforce compatibility and exchangeability, all the models in the repository are stored and extended according to the Predictive Model Markup Language (PMML) [5].

Finally, in the Respond stage of (iii) the developer specifies the Event-Condition-Action (ECA) rules to regulate the connection between the prediction results learned in the Predict stage of (ii) and the consequent action items, e.g., updating the monitoring logic with the safety threshold newly learned.

At run time, the working system will load the feature generation template. It will first check with the feature extraction rules and privacy control policies to determine which data fields to be included in an integrated view. Then it will treat uncertain data fields in the integrated view with the pre-specified data cleaning procedure, transform the integrated view into a feature set using the specified extraction rules, and finally load the feature set into memory. Before entering the prediction stage, the system will check whether the selected model exists or needs update. If a new model is required for prediction, it will be trained according to what is specified in the modeling schema. In the prediction stage, the loaded feature set will be validated against the selected model. Given the validation result, the system will determine the event and conditions and apply the ECA rules to trigger the right action in the response stage.

C. Scenario

When an end user logs on the Wellness Management Portal, the portal may remind the user to measure some vital signs. When the user initiates the process, the related service will apply the clinical guideline for vital signs measurement. When data are uploaded, the collected data are consumed by the Wellness Monitor Services. By executing the event processing logic, the monitor services examine the new received data with historic data to provide realtime feedback. The feedback can be an alert for the user to be aware of an out of range vital sign measurement result. In some case, notifications may be generated and delivered to wellness assistants for further investigation. The collected data and processing results are persisted in the Wellness Record & Knowledge Repository.

III. REALTIME WELLNESS MONITORING

Currently, realtime wellness monitoring focuses on treatment compliance management, for example, chronic disease care. Chronic disease treatment consists of clinical and home care. Home care is very important during the treatment, wherein most of the medical procedures are occurred. However, currently, there are not sufficient services that can facilitate home care. On the one hand, the end users lack of awareness on how well they adhere treatment plans. On the other hand, doctors either lack of information or tooling supports to understand the effectiveness of home care. In this section, we will present a solution that allows ISVs to develop new wellness monitor services that are able to provide better awareness.

A. System Requirements

Typically, monitoring on home care for end users focuses on vital signs, physical exercise, nutrition, etc. Due to the diversity of people’s wellness conditions, such as mobility, cardiac performance, and nutrition habits., the monitoring service requires personalization. Further, human’s activities and wellness progress are dynamic. End users may switch to different wellness services, or adopt new devices, etc. Also, the doctor may change treatment plans, according to progress of patients’ wellness recovery. This requires monitor services to support dynamicity as part of the non-functional requirements. Given the fact that the number of monitor services and end users may scale up, the platform needs to deal with scalability issues.

B. Cloud-based Wellness Monitoring Service

In our platform, we introduce a cloud-based wellness monitoring service framework. The framework provides: (i) Monitor service development API, for developing event processing logic for realtime feedback. The scope of the API is given in previous section. (ii) Service management API, for managing the dynamicity of the application logic. The API enables runtime evolution of event processing logic. The dynamicity of the application logic may include (a) change of event subscription and associated filtering predicates; (b) modification of metric definition and associated computation logics; (c) adjustment of alert generation logic. (iii) Elastic service runtime architecture. We propose a cloud-based infrastructure to hosting the monitoring services developed by ISVs. The infrastructure includes: (a) queue services for message transportation; (b) runtime engine for executing event processing logic; (c) runtime data store for persistence. It should be noted that these three components are self-management Platform-as-a-Service, i.e., they can automatically scale-out/in according to the workload.
IV. ANALYTICS FOR WELLNESS MANAGEMENT

Central to the development of a wellness management service is the analytic capability to transform wellness monitoring data into actionable insights. In this section, we use the scenarios of diabetic care to demonstrate the use of analytics and the associated challenges. In particular, we focus on three areas: disease management, lifestyle management, and disease learning.

A. Disease Management

In the first scenario, the process of transforming monitoring data into tailored feedbacks involves four major analytics tasks: (1) predict glucose concentration ahead-of-time; (2) determine abnormal glycaemic episodes; (3) regulate insulin delivery; and (4) generate personalized management plans. To address these tasks, we need to carefully craft models that describe the characteristics of the target phenomenon such that the models can be used to validate the conformity of incoming data for detection or ahead-of-time-prediction. For example, the monitor service needs to determine safe thresholds for glucose concentration and identify predictors of hyperglycemia and hypoglycemia episodes. A closed loop insulin delivery system also needs prediction models to determine optimal insulin delivery rates.

Despite the success of the physiology-based and data-driven modeling approaches in experiment settings (c.f. Section V), there still exist challenges in employing them to create analytics services. First, the creation of physiology-based compartmental models can be hindered by insufficient knowledge of the compartments, e.g., the structure and dynamics of the insulin hormone [6]. Inter- and intra-patient variability, which has been found in previous research as significant, has to be addressed each time the models are applied [7,8]. Second, the data-driven models also have to account for inter- and intra-patient variability before drawing predictions on the targeted patient. Moreover, the continuous glucose monitoring (CGM) scenario poses a new realtime analysis requirement to existing temporal modeling techniques.

In fact, there are several technical questions and research issues remain to be answered for the development of diabetes management service. The key questions include how to compensate for the uncertainty between the proposed models and targeted patients and, if necessary, how to adapt models for the patients. Specific to the CGM scenario, there are also questions about how to adapt the temporal models for the targeted patient and compute them efficiently. Lastly, since both the physiological and data-driven modeling approaches have shown potential in disease management, we can develop new approaches that integrate prior knowledge of the physiological model into the data-driven one.

B. Lifestyle Management

Previous studies have shown the effectiveness of lifestyle interventions in diabetes prevention and management. However, the computer-assisted programs have not been fully leveraged the awareness of the patient’s health conditions obtained by monitor services. For example, depending on the current glucose concentration, different types of diet plans and physical exercises can be proposed to the target patient to optimize glycemic control. Also, if the monitor service comes with the capability to recognize dietary intakes (e.g., by note-taking or a smart object recognition device) and physical exercise (e.g., by notes or by sensors), the information can also serve as pre-conditions for optimization.

This has led us to the challenge of casting the task of lifestyle intervention planning into an optimization problem. Previously, many frameworks were proposed to treat similar problems in the area of operational and statistical research, e.g., multi-attribute optimization and multi-attribute decision making. To apply these frameworks, we need to specify the knowledge of monitoring data and personal lifestyle history as optimization constraints. In addition, we also infer user models of the target patients based on their preferences and use the user models to adapt their lifestyle intervention plans.

C. Disease Learning

Monitoring data collected for disease and lifestyle management can also be accumulated to generate evidence for clinical decision support and disease learning. In general, there are at least three types of predictive diagnostic decisions: (1) prediction of the risk of developing diabetes; (2) early diagnosis of diabetes; and (3) prediction disease progression and related complications, e.g., cardiovascular diseases. Among these decisions, (2) and (3) are the most promising to be enhanced with addition information from the monitoring data. First, the monitoring data of the key glycemic control players (e.g., fasting plasma glucose, fasting serum insulin) will be collected and stored in the database. Then, the analytics component can perform predictive diagnosis on disease risk profiling (including progression and complications), leveraging the predictive models of the key risk factors.

Given the high risk of diabetes as well as individual predisposition to target organ injury, it is essential to develop such disease learning and predictive diagnostics applications to support clinical decisions on pre- and diabetes care. Despite the success of previous research in quantifying the properties of the likely disease evolution path with analytics models, there still exists challenge to apply these models in practice -- specifically, how to use the collected personal data to make accurate prediction of disease progression and complications on the individual level for clinical decision support. As for the research of disease learning, there also remains the question about how to find similar patients from the pool of accumulated monitoring data. Automatic (and semi-supervised) clustering is applied to identify sub-population that has distinctive sensitivity to certain external factors and propensity to certain complications.

As we discussed earlier, analytic services provides programming models and related APIs, testing/debugging data sets, and a runtime environment that facilities ISVs to develop new applications for the above three areas. It should be noted that knowledge generated by the analytic applications can also be deployed as event processing logic to assess wellness status in realtime fashion.
V. RELATED WORK

A common expectation among the research community is that monitoring data can serve to improve the different stages of the diabetes management process. For instance, the analytics component improves the self-managed insulin delivery system by providing instant glycemic control suggestions and generating alerts of hypo-/hyper-glycemia conditions to patients and their caregivers. Furthermore, with the recent advances in subcutaneous and non-invasive optical sensors, there comes the need of a real-time analytics component that can transform the CGM data into useful information, e.g., prediction of next-hour change of blood glucose concentration. Clinical studies have shown that CGM is effective for maintaining glycated hemoglobin at a safe level and preventing severe hypoglycemia episodes [9]. In the cases that a closed-loop automatic control system is adopted, it is imperative to adjust the dose and timing of insulin injection with respect to the anticipated changes.

Diabetes researchers have proposed different ways to learn analytics models based on the different types of monitoring data. Some employ the understanding of metabolic dynamics in diabetes patients to develop physiological-based compartmental models. The models are then used to simulate glucose-insulin interaction and to trigger proper responses for preventing hyperglycemic episodes [8,10]. Other researchers in bioengineering attempt to learn predictive models from data directly. For example, some infer statistical models to learn how glucose concentration changes in response to external factors such as diet, physical activity and medication [11]. Recently, with the CGM data available, more and more researchers focus on inferring models from time series to predict near-future glucose concentration [12,13].

Wellness data has also been expected to lend support to the creation of other wellness management services. In fact, many large-scale studies have confirmed the effectiveness of life-style interventions, such as modification of dietary intakes and increase of physical activity, on the prevention and treatment of diabetes [14,15]. This is particularly important for patients who have been diagnosed of non-insulin dependent diabetes mellitus (Type II Diabetes).

Lifestyle interventions are traditionally promoted through patient education. Recently, there has been discussion about building computer-assisted programs to provide frequent feedbacks and tailored plans of behavioral modification. Controlled random trials have presented evidence on its effectiveness in improving patient adherence and clinical outcomes [16].

Finally, wellness data collected for disease and lifestyle management can be accumulated to generate evidence for clinical decision support and disease learning. They can help identify leading indicators of disease progression in different cohorts [17,18]. Previous studies have also developed physiological models to simulate the interrelationships of the key players in the control system [19].

VI. CONCLUSION

In this paper, we advocate an open platform for wellness management ecosystem. The key components and related services of the platform are illustrated. In particular, wellness monitoring and analytic services are presented. By providing open API and service management facilities, we expect ISVs to rapidly develop/deploy new services and related business model by focusing on core business logics such that an ecosystem can be formed for wellness management. Such an ecosystem is also expected to facilitate the development of wellness management services and applications, resulting in high quality and low cost healthcare systems. We will revisit these expectations in the near future when the platform is deployed in a real-life setting.

REFERENCES