Journal of Universal Computer Science, vol. 24, no. 9 (2018), 1153-1169
submitted: 30/12/17, accepted: 30/6/18, appeared: 28/9/18 © *J.UCS*

Longitudinal Healthcare Data Management Platform of Healthcare IoT Devices for Personalized Services

Ahyoung Choi
(Gachon University, Seongnam, Gyungki, South Korea
aychoi@gachon.ac.kr)

Hangsik Shin[1]
(Chonnam National University, Yeosu, South Korea
hangsik.shin@jnu.ac.kr)

Abstract: Recently, many studies have been conducted on how to manage and analyze various types of health data such as clinical data, genomic data, and wirelessly collected multiple sensory data. In this paper, we propose a web-based healthcare data integration and management platform that collects heterogeneous types of health-related medical record as well as real-time lifelogging data. This platform provides flexible architecture to different types of data exchanges. The platform manages real-time data such as heart rate, blood pressure, and activity information extracted from various healthcare devices and provides functions to transmit them to the server. Then it analyses the risk based on a domain knowledge and individual differences by applying machine learning tools, then visualizes the result to the patient and doctor dynamically based on information simplification method. It also controls the data access authority concerning the level of expertise and role. For evaluation of integrated data analysis, we apply open database and evaluate the proposed risk analyser result. The proposed platform could be utilized for future healthcare service to share accumulated healthcare data in various situations.

Keywords: Healthcare data; lifelogging data; electrical medical record; data integration; data management
Categories: H.3.4, H.3.5, H.5.1, H.5.2, E.0

1 Introduction

Big data and IoT research in the field of healthcare and wellness care have been actively carried out in recent years [Laney 2001, Hopkins et al. 2011, Chen et al. 2014, Hu et al. 2014, Belle et al. 2015, Raghupathi et al. 2014, He et al. 2017, Shameer et al. 2017, Santos et al. 2016]. This has largely been because of an increase in both the number and type of digitized data; as a result, heath-related big data processing and management methods is an active area of research. Digitized medical data includes clinical data, such as medical imaging data, prescription records, and lab data, as well as personal health data (PHR) such as personal profile, obesity level and family medical history. In addition, many researchers have mentioned precision medicine as a model for the next generation of healthcare [Chen et al. 2016]. Precision medicine is an emerging approach to focus on prevention and

[1] Corresponding author

treatment considering individual differences in genes and lifestyles rather than focusing on diagnosing the disease in a uniform way by population-based statistics and science. This approach accelerates the health monitoring and wellness monitoring device market and its integration of traditional medical records with the viewpoint of big data. For example, GearFit, Fitbit, Jawbone, G Watch Urbane and other wearable devices monitor daily exercise activity, sleep quality, and dietary habits for wellness monitoring. Portable blood pressure monitors and patch-type ECG measurement sensors collect health-related data in an unobtrusive way in daily life for health monitoring.

With this trend, it becomes important to integrate and analyse data measured from various heterogeneous wellness and health monitoring devices. However, access to such medical records for specific purposes is difficult because hospitals and insurance companies own most medical records. In addition, data formats vary from hospital to hospital, therefore it takes considerable effort to integrate information from multiple institutions. Moreover, various medical imaging data in hospitals, such as fMRI, X-Ray, as well as vital sign data are measured in real time and are managed by hospital departments; thus various types of data are not integrated and intelligently analyzed. To address these problems, many researchers have proposed lifelong and patient-owned health data gathering platform and its data exchange method [Sunyaev et al. 2010, Masud et al. 2015]. The Microsoft Health Vault is a web-based data management platform designed to allow patients to manage their health information directly [Sunyaev 2010]. It provides functions to connect more than 180 health monitoring devices such as blood pressure monitor, glucose meter, digital scales and 25 applications such as Moves, Preva etc. However, it does not integrate and examine EMR data and real-time monitoring data. Medtronic ZephyrLIFE is a remote patient monitoring application that manages heart rate, respiratory changes, electrocardiogram (ECG) by attaching bio patch to the patient's body. Many researchers have also studied remote patient monitoring platforms for data integration and management [Wang et al. 2014, Liang et al. 2012]. In their application, data is extracted from existing wearable devices or portable medical devices and users in real-time. However, there was a lack of functions to present the status of the patients and there is no description about longitudinal data processing and the analysis of personalized healthcare and wellbeing care.

From related works, there exit many issues to be concerned with health data analytics platform: first, the platform should include heterogeneous data that are real-time data as well as non-real-time data. Thus, it is required a seamless transparent mechanism to access different types of data such as EMR systems stored in a relational database, PHR systems stored noSQL database by following DICOM/HL7 standard medical data transfer protocol as well as sensory data obtained from healthcare IoT devices. Therefore, the novel platform needs to be extended to various types of data sources through flexible architecture and data exchange methods. Secondly, after understanding different types of data, we need to access it and analyse concerning a domain knowledge and individual differences. It is required to support an automatic analysis by various kinds of machine learning tools. In addition, appropriate visualization depending on level of expertise in a specific domain need to be considered. The degree and method of information to be appropriately provided depends on the user's expertise to build a general platform. It is important that the abstraction of this information be performed automatically, not manually. Thirdly, privacy and data access issues need to be considered. For one-person data, many stakeholders such as insurance company, family and doctors are related, and conflicting interest may occur among stakeholders. Thus, data access management considering privacy issues need to be considered. If users manage the accessibility and authority of the data manually, it can be a burden to use the system. Therefore, the platform should consider automatic management of data considering the privacy of the data provider and the role and expertise of the person who wants to use the data. Lastly, it manages longitudinal data for personalized health and well-being care application.

In this paper, we propose a web-based healthcare data integration and management platform that collects heterogeneous types of health-related medical record as well as real-time lifelogging data. This platform provides flexible architecture to different types of data exchanges and it analyses the risk based on a domain knowledge and individual differences by applying machine learning tools, then visualizes the result to the patient and doctor dynamically based on information simplification method. It also controls the data access authority concerning the level of expertise and role. We evaluate the proposed risk analysis functions using a public database, MIMIC, which was used to collect multiple parameter data of ICU patients for 24 hours monitoring [Moody et al. 1996, Goldberger et al. 2000]. As a result, we conclude that the proposed data management platform could be utilized for remote medical service or healthcare home service to provide a function for sharing and confirming accumulated medical data in various situations.

2 Related works

In these days, research is actively being conducted on big data processing and analysis in the fields of healthcare and well-being care. Big data is described as a 3Vs model (or a 4Vs model) that includes the following features: volume (i.e., amount of data), velocity (i.e., speed of data processing), variety (i.e., types of data) and, assuming 4Vs, veracity (i.e., accuracy of data) [Chang et al. 2016]. With regards to a large volume of data, Chang assumed that the average size of electronic medical records were 20 megabytes per image and estimated the total size of the big data in the medical field to be 100-250 exabytes with an annual average growth rate of 1.2-2.4 exabytes [Sun et al. 2013]. Researchers at IBM reported more than 1 trillion obtained from connected objects and devices at 2010 and researchers at Simens estimated data explosion by 2020 up to 40 zettabytes [Raven et al. 2015]. Therefore, it has become important to develop a scalable platform to handle a large volume of medical data.

Considering increased volume of data, many researchers have applied the previous big data management method in the health sector. Hussain and his colleagues proposed mobile sensory data analysis method using MapRedue in cloud computing environment [Hussain et al. 2014]. Wang and his colleagues proposed the mathematical model to remove duplicate data in a big data storage system [Wang et al. 2016]. To help understand the various medical terminology, Can and his colleagues developed relational and ontological triple stores for healthcare domain knowledge sharing [Can et al. 2017]. However, access to such medical records for research purposes limited due to the protection of personal information. Hospitals and insurance companies owned most medical records and it is difficult to collect and utilize medical data outside of the hospital. Therefore, storing medical records in cloud platform might impractical solution at this stage unless it is regulated or legally supported. That is, ultimately, data analysis and processing storage are expected to take place in the cloud platform, but now medical data such as EMR, medical imaging data and so on is limited in that access to the data by individual privacy protection regulations and rules. In the web-based platform, it is a realistic alternative because it is an open access platform to share and utilize patient data when the patient agrees on utilizing his or her data.

Another issue of big data research in the medical field are considered to be the integration of many types of digitized medical data. Digitized medical data includes clinical data such as EMR (Electronic Medical Records) or EHR (Electronic Health Records), personal health data (PHR) and physiological sensory data obtained from wellness and health monitoring devices. Clinical data includes medical imaging data, prescription records, lab data and genomic sequence data obtained from various medical centers. Personal health data (PHR) include personal profile, obesity level and family medical history and so on. Wirelessly collected multiple sensory data such as heart rate, body temperature and blood glucose levels collected by wearable devices are another type of health and wellness data. In particular, GearFit, Fitbit, Jawbone, G Watch Urbane and other wearable devices monitored daily exercise activity, sleep quality and dietary habit for wellness monitoring. For health monitoring, portable blood pressure monitors, and patch-type ECG measurement sensors called AliveCor collect health-related data in a daily life. However, integration of various health related data requires lots of effort since data formats vary from hospital to hospital as well as wellness data is not formatted by medical standard.

With this trend, many companies and researchers have proposed data management platforms to integrate data measured from various heterogeneous monitoring devices. The Microsoft Health Vault is a web-based data management platform designed to allow patients to manage their health information directly [Sunyaev et al. 2010]. It provides functions to connect more than 180 health monitoring devices such as blood pressure monitor, glucose meter, digital scales and 25 applications such as Moves, Preva and so on. Additionally, it provides data management functions that individual users upload their EMR data and real-time monitoring data. However, it does not integrate and analyze EMR data and real-time monitoring data. In addition, it supports to collect the data of predefined devices only which listed on the website. Apple HealthKit and ResearchKit provide a method to collect health, fitness and medical research data from wearable and mobile devices. Medtronic ZephyrLIFE is a remote patient monitoring application that manages heart rate, respiratory changes, electrocardiogram (ECG) by attaching bio patch to the patient's body. The system receives the signal from various kinds of devices and analyzes based on Health-Hub platform in the home and the hospital. Recently they provide stress analysis and exercise performance analysis using their wearable chest type sensors. Liang et al. proposed a method of collecting heterogeneous sensor information for home health care service. In their application, data was extracted from existing wearable devices or portable medical devices and users were monitored in real-time [Liang et al. 2012]. Khalid Elgazzar et al. developed a ubiquitous health monitoring system using mobile web services to support mobility of chronic disease patients [Elgazzar et al. 2012]. If the patient registered through the website, vital signs were monitored and transferred to the doctor side. However, most previous works have recommended that the platform should integrate multiple sensory data or clinical data for remote patient monitoring purposes. There is still not much research on longitudinal data processing and analysis for personalized healthcare.

After integrating and managing various types of data, it is required to estimate users or patient status based on the collected data. Typically, clinical decision support system (CDSS) was used for diagnosis and prevention of disease. The existing CDSS uses a method in which doctors or medical experts explicitly input domain knowledge in a rule-based engine. However, recently, IBM's Watson has been developed to automatically predict and judge disease without explicit inputs by utilizing machine learning technology. Many researchers have also studied health data analytics methods by data integration and management. Muhanmmad Bilal Amin proposed a health and wellness platform to process cloud-based multimodal sensory data. In this platform, raw sensory data is continuously collected in real time regardless of the kinds of devices and the lifelogging data of the user is extracted based on the context. Intelligent monitoring based on an expert-driven rule is made possible by Lifelog extraction [Amin et al. 2016]. Especially, in raw sensory data acquisition, a timeframe-based synchronization method is applied by using complete-sync, incomplete-sync mode via sensory data queue to synchronize data obtained from various sources. Luo developed a tool to estimate the best machine learning algorithms and parameters for big clinical data analysis [Luo et al. 2016]. Adikaram proposed continuous learning method based on graphical knowledge using of highly density big data [Adikaram et al. 2016]. In addition, an appropriate visualization method was also proposed [Seo et al. 2015]. However, those previous works lack of function to analyze and present the status of the individual users. Another important issues regarding data acquisition and management for distributed health care systems are privacy and security. Many researchers have been proposed role-based access control based on the information of user, roles, permission, and sessions [Sandhu 1996, Bertino 2001]. However, data access of those models is limited in dynamically changing environments. To handle arbitrary and dynamic authorization polices, context based access control method have been proposed [Hu 2004, Zhang 2004, Ray 2016]. Those models provide permission on data based on the context information not only role but also location, time, and other data access condition. These research on context-based access control of distributed health related data, but not much research on access control of mobile health data such as personal health record, dynamic sensory information such as mobile ECG monitoring devices.

From related works, there exit many issues to be concerned with health data analytics platform: first, the platform should include heterogeneous data that are real-time data as well as non-real-time data. Thus, it is required a seamless transparent mechanism to access different types of data such as EMR systems stored in a relational database, PHR systems stored noSQL database by following DICOM/HL7 standard medical data transfer protocol as well as sensory data obtained from healthcare IoT devices. Therefore, the novel platform needs to be extended to various types of data sources through flexible architecture and data exchange methods.

Secondly, after understanding different types of data, we need to access it and analyse concerning a domain knowledge and individual differences. Thus, it is required to support an automatic analysis by various kinds of machine learning tools. In addition, appropriate visualization depending on level of expertise in a specific domain need to be considered. The degree and method of information to be appropriately provided depends on the user's expertise to build a general platform. For a group of specialists such as physicians, it is more effective for the patient to understand the measured information and the resource itself, and for the user or the care giver to present the information at the abstract level to understand the meaning of the information. It is important that the abstraction of this information be performed automatically, not manually. Thirdly, privacy and data access issues need to be considered. For one-person data, many stakeholders such as insurance company, family and doctors are related, and conflicting interest may occur among stakeholders. Thus, data access management considering privacy issues need to be considered. If users manage the accessibility and authority of the data manually, it can be a burden to use the system. Therefore, the platform should consider automatic management of data considering the privacy of the data provider and the role and expertise of the person who wants to use the data.

3 Lifelogging data platform of healthcare IoT devices

3.1 Overall architecture

The web-based platform for the integration of the proposed real-time and non-real-time medical data is structured as shown in Figure 1. The data integration platform consists of a data layer, a process layer, and a service layer. Data layer provides a function to acquire heterogeneous types of health-related medical record as well as real-time lifelogging data. In this step, medical record manager translates the data by following medical standard format and sensor manager uses sensor API to obtain sensory data. Sensor manager collects a lifelogging data, such as heart rate, blood pressure, GPS, and activity information extracted from various healthcare devices and transmits them to the server through wireless communication. Depending on sensor data type, sampling frequency, data description and format is totally different. Therefore, sensor manager interpolates the missing data and removes replicated data according to type of sensors. Non-real-time medical data, such as height, weight, age and gender, and lab test information measured by a medical institution are transmitted to the server in XML format. For medical record data, electronic medical record represented by HIS/RIS and medical imaging format represented by PACS protocol should be integrated and understood by checking the header format. Medical record data gateway integrates data having different standards, and is standardized in the form of XML and stored in the repository. In case of sensor data, streaming data is integrated through PHR and sensor data gateway, and is stored in the repository together with time information and information about the measured sensor.

After identifying and collecting the heterogeneous data, data aggregation, semantic data integration and personalized risk analysis performed in a process layer. Data management module manages health data repository and device or user profiles. It checks novel update on repositories and stored up-to-date medical record and sensory data. Data aggregation module collects the steady-state data and real-time data, and semantic data integration interprets and analyse current health conditions based on health ontology. Health ontology may cover several CDSS rules about irregular body conditions. In data integration, the collected data is integrated based on the time information. When the same information is collected from other sensors, priorities are determined based on known accuracy information and data precision. Then, a personalized risk analyser function provides a function that simplifies the event through the crisis analysis data for a lifetime. The risk analyser managed the user's health status based on integrated real-time, non-real-time health data and provided feedback through the web server. This platform analyses the risk based on intelligent analysis methods such as machine learning tools, then visualizes the result to the patient and doctor dynamically.

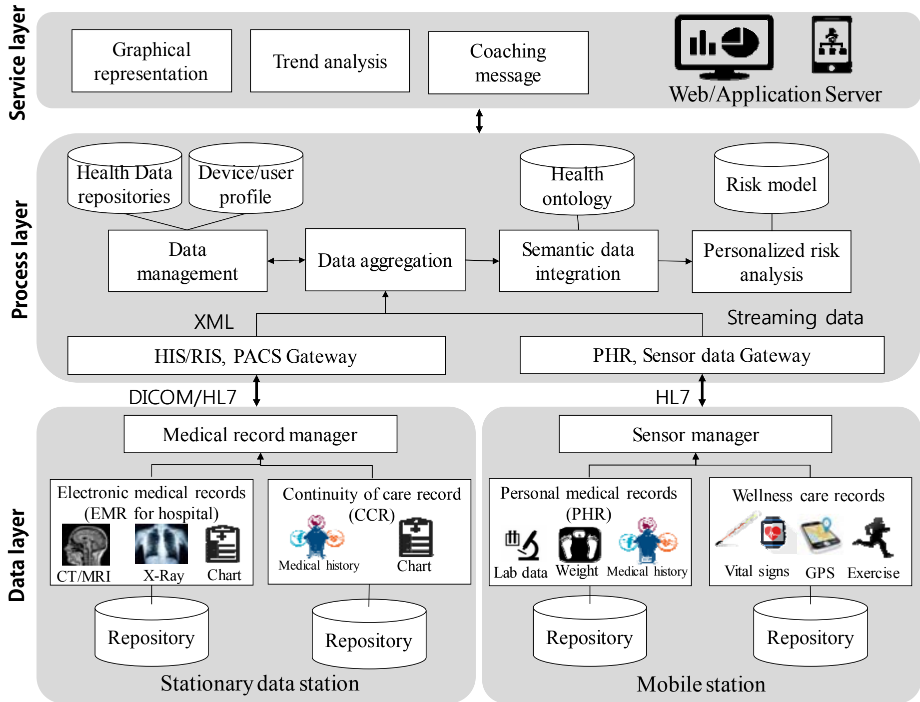


Figure 1: Overall architecture

The web browser for personalized healthcare service provides feedback through the web server in service layer. In services layer, it is important to visualize healthcare information appropriately based on on level of expertise in a specific domain. The degree and method of information to be appropriately provided depends on the user's expertise to build a platform. For a group of specialists such as physicians, it is more effective for the patient to understand the measured information and the resource itself, and for the user or the care giver to present the information at the abstract level to understand the meaning of the information. It is important that the abstraction of this information be performed automatically. This information visualizes by graphical representation, trend analysis, and coaching message generation. To simplify the information, we design two different kinds of visualization pattern, one is for expert, the other is for non-expert user. Privacy and data access issues are considered concerning stakeholders. The platform supports automatic management of data considering the privacy of the data provider and the role and expertise of the person who wants to use the data. To tackle this issue, we proposed context based access control method to deal with dynamic configurations of heterogeneous health data.

3.2 Data repository design for heterogeneous data and access control

For the integration of real-time medical data and non-real-time medical data, a sustainable database was constructed as shown in Figure 2. The database constitutes the electronic medical record (EMR) data, the personal health related (PHR) information record and a patient-doctor relationship record for a remote patient monitoring service. The medical-related data followed an XML format because most classical medical data supported XML file format. To manage EMR data, users upload their medical record by XML file format and the system automatically parses it and stores in the database. In this study, it is assumed that the basic medical record is acquired in the form of XML, but it will be updated to support the EMR protocol conforming to the international protocol or to support various data types as a future work. To manage PHR data, PHR related database stores a user's personal profile by manual input and real-time lifelogging data obtained from wearable devices and medical monitoring devices by the sensor manager module of the mobile station. An example of PHR data exchange protocol is shown in Table 1.

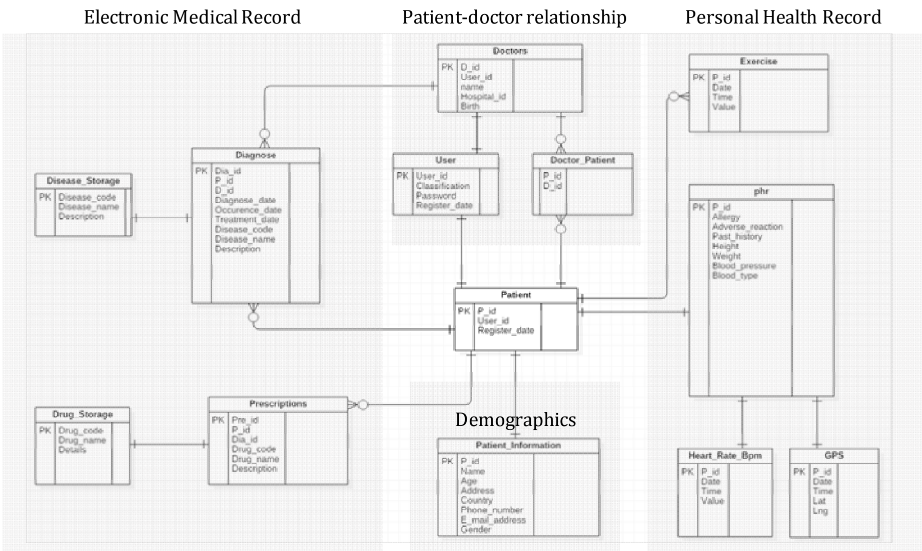


Figure 2: Data schema

Category	Type	Field	Category	Type	Field
User profile	UserID	110191812	Vitalsign	Type	bloodpressure
	Age	80		Date	20171230
	Gender	Female		Value	120/80

Table 1: Example of sensory data protocol

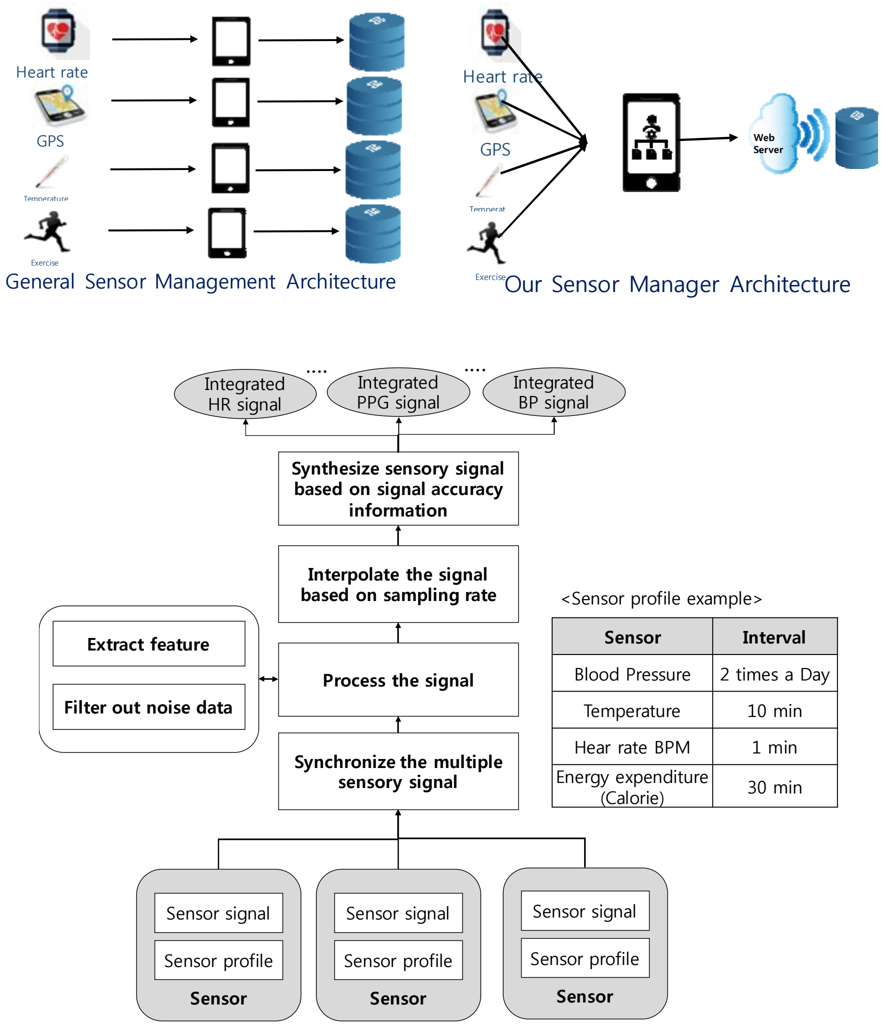
Data access control consists of three main process; login, check context condition and show information by role of user. If user ID and password is correct, user can login on application. Then, user wants to access certain patient information, application checks context condition based on Table 2. An application checks primary constraint context automatically such as time and location and secondary context such as role and expertise level. If user satisfies the primary context, user can access to information of patient. Then the system checks secondary context to confirm the level of data access. 'Role' attributes include 'Doctor', 'Nurse', 'Researcher', 'Family' and 'Others'. It determines what information is displayed to the users. And 'Expertise' attributes include 'Primary', 'Secondary' and 'Reference'. It determines how much information is disclosed and what kinds of action is permitted. If a user has primary attribute, user can read, edit and delete information. Otherwise, if a user's attribute is secondary, user can only read information. From this process, the types of data that can be accessed and the actions that can be taken are determined automatically. For example, 'Doctor' only have location context, 'Researcher' has Location context and Time context between 11:00 AM and 6:00 PM. When user is a doctor and primary user. user can access all the data of patient and full operability to manipulate the data such as reading, editing and deleting the information. If the user is a researcher and secondary user. He only can see limited PHR of patient, and he can read and edit the information but can't see or edit other data. Table 2 shows the information access result based on access user type. Third example is a case that access request user is a family with patient. He can access to limited PHR and EMR with read-only option.

Scenario	Primary Context (Time, location)	Secondary Context (Role, Expertise)	Result
#1 Doctor	(Location == hospital)	(Role == 'Doctor') ∩ (Expertise == 'Primary')	Activity: Read, Edit, Delete Data: PHR, EHR, EIR
#2 Researcher	(Time >= 11:00) ∩ (Time <= 18:00) ∩ (Location == hospital)	(Role == 'Researcher') ∩ (Expertise == 'Secondary')	Activity: Read, Edit Data: limited PHR
#3 Family	(Time >= 9:00) ∩ (Time <= 19:00) ∩ (Location == hospital)	(Role == 'Family') ∩ (Expertise == 'Reference')	Activity: Read Data: limited PHR, limited EHR

Table 2: Example of data access control

3.3 Data integration, analysis, and visualization

The sensor manager, medical record manager, data aggregation, and data integration module process the data to support health data integration and analysis. Sensor manager provides functions to manage multiple real-time health related sensory data as shown in Figure 3. It controls data collection interval, processes to filter and feature extraction based on sensor profile information. Then it sends data to server for furture integration. Sensor data includes sensor profile (type, sampling rate), sensor signal and signal property (accuracy, granularity).



Sensor	Interval
Blood Pressure	2 times a Day
Temperature	10 min
Hear rate BPM	1 min
Energy expenditure (Calorie)	30 min

Figure 3: sensor manager architecture and process

Clinical records such as EMR, EHR, and PHR data are transferred to the medical record manager. The medical record manager reads different types of files and sends them in the form of XML and any other file types that data aggregation module can understand. The sensor manager is designed to collect and manage various kinds of lifelogging data. Smart wearables connected to the mobile device and sensor data was received by wireless communication. Since the data transmission time and the cycle may be different, depending on the type of data, the transmission period can be adjusted by a user's input. Once the characteristics and cycles of the data are set, information is then converted into XML data and transmitted to the platform server. Data aggregation module checks header name and runs resampling if the collected data has a different range and sampling rate. In the data integration part, the collected data is integrated based on the time information. When the same information is collected from other sensors, priorities are determined based on known accuracy information and data precision.

In a risk analyzer, the possible risk is analyzed based on the conditions of the patients with heart disease and the stored EMR data and the PHR data. A visualization technique is applied so that the user recognizes the data change easily. For risk assessment, we specified the target users, especially for patients with heart diseases. Heart rate, blood pressure, GPS, and step count data were received for this application. The risk outcome is derived based on the currently measured heart rate, blood pressure, exercise intensity and personal profile such as sex and age. It detects normal heart rate range and determines whether the current user has an emergency health condition. In addition, blood pressure and weight, which are related to heart disease, are also analyzed by monitoring the risks. Machine learning technology such as SVN, linear regression, and Neural Network method is applied to analyze the data gathered.

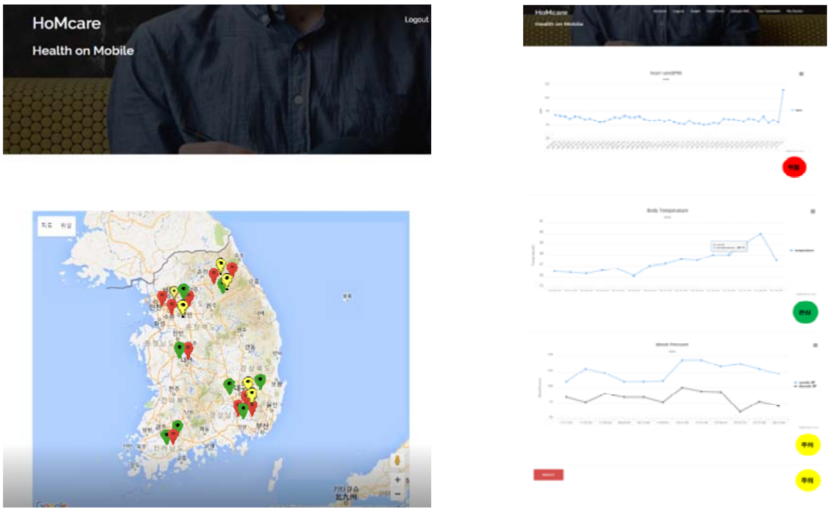


Figure 4: Simplified data visualization result

After analysing the data, the system visualize the information based on user's context such as expertise level, access purpose and so on. To support these function in an automatic manner, we define information visualization step as follows. Based on the result of information disclosure in access control session, the platform determines what kinds of data such as medical imaging data, electrical medical records and so on will be disclosed. It is effective for the patient to understand the measured information and the resource itself for a group of specialists such as physicians. In addition to this, the researchers or the care giver also can review the information at the abstract level to understand the meaning of the information. If the signal changes more than a certain level compared to the normal level or it is classified as an abnormal condition; the risk analyzer then relays the result to the user by visualizing a simple light sign metaphor. The current emergency level expressed in red, yellow, and blue colors. Emergency level is obtained by averaging each feature's risk level. Feature's risk level is determined by the standard guidelines. For example, normal heart rate range is known from 60bpm to 100bpm. Algorithms for risk analysis depend on the target application.

4 Implementation and experimental results

To establish the platform, we did develop initial prototype version. Many previous frameworks provide conceptual architecture and concepts for data exchanges. However, in this work, we developed main part of each layer and showed the effectiveness of the proposed method. For this purpose, this study was carried out with the case of patients with heart related diseases as an example. We acquired EMR data from patients, measured heart rate in real time to integrate data, and constructed a platform to analyse personalized risk considering ontology of heart related disease and current condition of patient. The implementation results of the web-based data integration platform are as shown in Figure 3 and Figure 4. The server was built using Linux, the DB built on MySQL and NoSQL and the LG G Watch Urbane was used to acquire real-time sensor information. In addition, the web was constructed using WordPress to use the service of the user and the user directly stored and managed the PHR and EMR data through the web. The current data was used based on scenarios for patients with heart disease, so that only heart rate, GPS, and step count data were received. The frequency was at 60 Hz for the heart rate, 1 hour for the GPS and 30 minutes for the step counter. Sensor manager was developed based on Android platform.

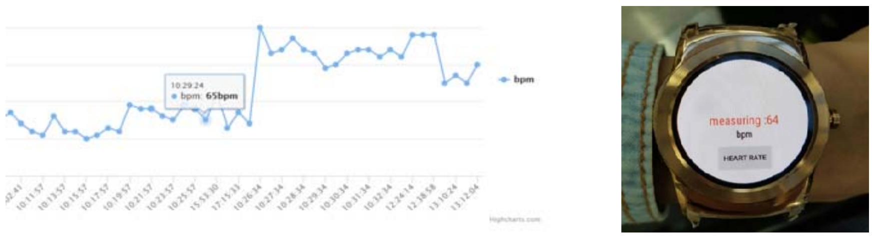
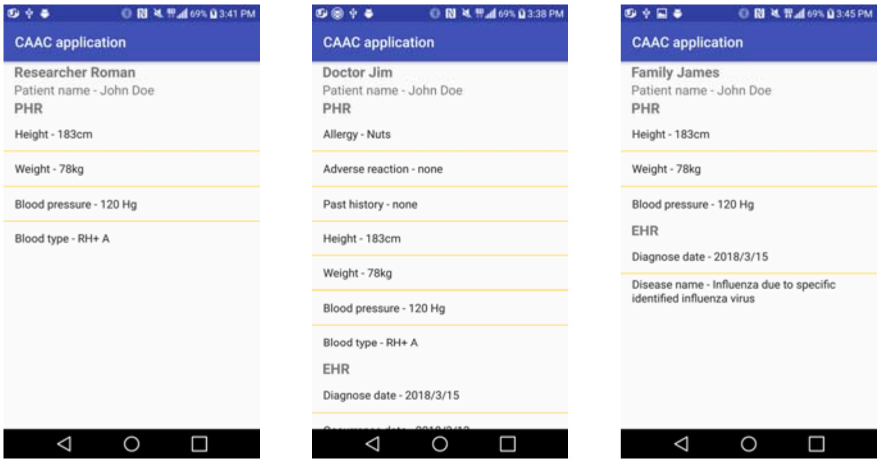


Figure 3: real-time data collection with wearable heart monitoring sensor



(a) (b) (c)

Figure 4: Data access control result (a) researcher (b) doctor, (c) family

The proposed data management platform can be utilized for remote medical service or healthcare home service to provide a function to share and confirm accumulated medical data in various situations. We evaluated the proposed risk analysis functions using a public database, MIMIC as an example of big data, which collected multiple parameter data of ICU patients for 24 hours monitoring as shown in Figure 6 [35, 36]. In the MIMIC database with a total size of 3.1 GB, multiple physiological signals were used for evaluation. Here we applied continuous ECG data and systolic blood pressure to classify the level of risk. For pre-processing, all signals were divided by period and excluded from analysis when one of the signals included an error.

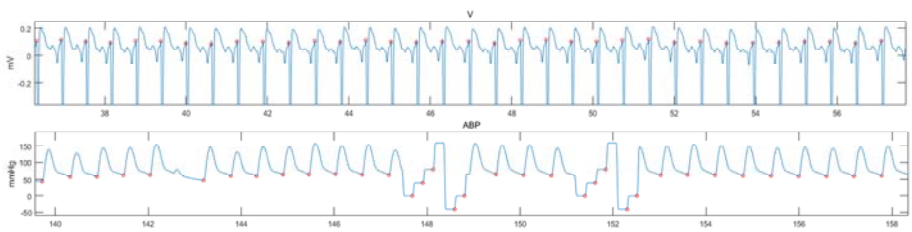
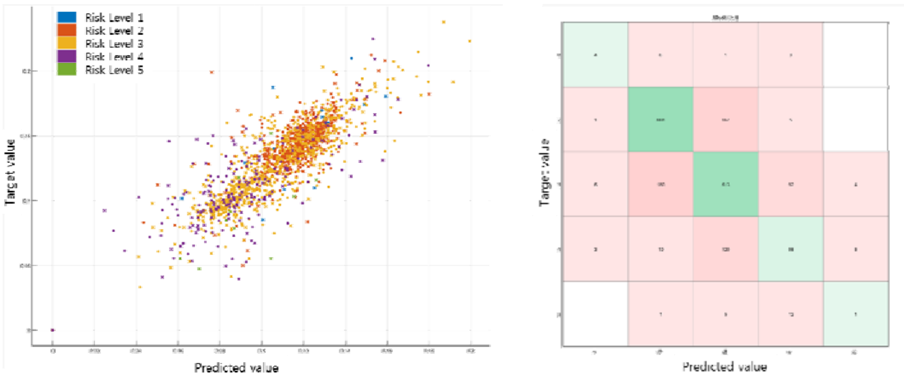


Figure 5: Sample data of MIMIC database; Uppercase (ECG), Lowercase (Systolic blood pressure)

For training, MIMIC database was used. For evaluation, PPG signal obtained from sensing devices and MIMIC database were used after integration. To evaluate, PPG signal is synchronized by ECG peak and inputted as an input of evaluation. As a result, a five-scale risk level was estimated with 90.3% accuracy, compared to reference data when Medium Gaussian SVM was applied. A threefold cross-validation was applied to a total 3500 data samples as shown in Figure 6. In this paper, we show an example of data analysis using SVN. Instead of SVN, various deep learning or machine learning algorithms provided by open API can be applied. At this stage, handling big data in real-time was not considered. We used local workstation to collect and to process the data using GPU processor to see feasibility of the proposed platform. As a future work, real-time processing of big data could be considered in near future.



(a) (b)

Figure 6: Risk analysis result: (a) Correlation plot; (b) Confusion matrix

5 Conclusion and future work

In this paper, we proposed a web-based healthcare data integration and management platform that collects heterogeneous types of health-related medical record as well as real-time lifelogging data. This platform provided flexible architecture to different types of data exchanges. The platform manages real-time data such as heart rate, blood pressure, and activity information extracted from various healthcare devices and provides functions to transmit them to the server. Then it analyzed the risk based on a domain knowledge and individual differences by applying machine learning tools, then visualizes the result to the patient and doctor dynamically based on information simplification method. As a result, we implemented that the life-logging data acquired in real time and the non-real-time medical data were periodically updated in the database and visualized through a web browser. For evaluation of integrated data analysis, we apply open database and evaluate the proposed risk analyser result. As a future work, we will collect diverse database that follows different types of medical standards and will extend the platform to support various kinds of heterogeneous sensors that operate on different platforms such as PPG sensors for an Arduino platform. In addition, we extend the various types of ontologies for semantic integration not only for heart related condition monitoring.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2016R1D1A1B03935104) and was supported by the Bio & Medical Technology Development Program of the NRF funded by the Korean government, MSIP (NRF-2016M3A9F1941328). We thank Yoomin Kim, Hansol Kim, Junghan Ji, Chiyoon Song for obtaining and processing the data.

References

[Adikaram, 16] Adikaram, K., Hussein, A., Effenberger, M. and Becker, T.: Continuous Learning Graphical Knowledge Unit for Cluster Identification in High-Density Data Sets. Symmetry 2016, 8(152), 1-17, doi:10.3390/sym8120152.

[Amin, 16] Amin, M., Banos, O., Khan, W., Bilal, H., Gong, I., Bui, D., Cho, S., Hussain, S., Ali, T., Akhtar, U., Chung, T. and Lee, S.: On Curating Multimodal Sensory Data for Health and Wellness Platforms. Sensors 2016, 16, 980, doi:10.3390/s16070980.

[Belle, 15] Belle, A., Thiagarajan, R., Soroushmehr, S. M., Navidi, F., Beard, D. A., and Najarian, K.: Big data analytics in healthcare. BioMed research international 2015, 2, 1-16, doi:10.1155/2015/370194.

[Bertino, 01] Bertino, E., Bonatti, P. and Ferrari, E.: TRBAC: A Temporal Role-Based Access Control Model. ACM Transactions on Information and System Security 2001, 4(3), 191-223.

[Can, 17] Can, O., Sezer, E., Bursa, O and Unalir, M.: Comparing Relational and Ontological Triple Stores in Healthcare Domain. Entropy 2017, 19, 30, doi:10.3390/e19010030.

[Chang, 16] Chang, A.: Big data in medicine: The upcoming artificial intelligence. Progress in Pediatric Cardiology 2016, 43, 91–94, doi: doi.org/10.1016/j.ppedcard.2016.08.021.

[Chen, 14] Chen, M., Mao, S. and Liu, Y.: Big Data: A Survey, Mobile New Appl 2014, 19, 171-209, DOI 10.1007/s11036-013-0489-0

[Chen, 16] Chen, Y., Guzauskas, G., Gu, G., Wang, B., Furnback, W., Xie, G., Dong, P. and Garrison, L.: Precision Health Economics and Outcomes Research to Support Precision Medicine: Big Data Meets Patient Heterogeneity on the Road to Value. J. Pers. Med. 2016, 6, 20, doi:10.3390/jpm6040020.

[Elgazzar, 12] Elgazzar, K., Aboelfotoh, M., Martin, P., Hassanein, H.S.: Ubiquitous Health Monitoring Using Mobile Web Services. Procedia Computer Science 2012, 10, 332-339.

[Fritz, 14] Fritz, T., Huang, E., Murphy, G. and Zimmermann, T.: Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Canada, April 26-May 01, ACM New York, NY, USA, 2014. pp. 487-496.

[Goldberger, 00] Goldberger, A.L., Amaral, L., Glass, L., Hausdorff, J.M., Ivanov, P., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000, vol. 101, no. 23:e215-e220.

[He, 17] He, K., Ge, D. and He, M. Big Data Analytics for Genomic Medicine. Int. J. Mol. Sci. 2017, 18, 412, doi:10.3390/ijms18020412.

[Hopkins, 11] Hopkins, B., Evelson, B.: Expand Your Digital Horizon With Big Data, Forrester Research Inc., 2011

[Hu, 04] Hu, J. and Weaver, A. C.: A Dynamic, Context-Aware Security Infrastructure for Distributed Healthcare Applications. In Proceedings of the PSPT 2004, 1-8.

[Hu, 14] Hu, H., Wen, Y., Chua, T. and Li, X. Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. IEEE Access 2014, 2, 652-987, doi:10.1109/ACCESS.2014.2332453.

[Hussain, 14] Hussain, S., Bang, J., Han, M., Ahmed, M., Amin, M., Lee, S., Nugent, C., McClean, S., Scotney, B. and Parr, G.: Behavior Life Style Analysis for Mobile Sensory Data in Cloud Computing through MapReduce. Sensors 2014, 14, 22001-22020, doi:10.3390/s141122001.

[Kaewkannate, 16] Kaewkannate, K. and Kim, S.: Comparison of wearable fitness devices. BMC Public Health 2016, 16(433), doi: 10.1186/s12889-016-3059-0.

[Laney, 01] Laney, D.: 3-d data management: controlling data volume, velocity, and variety, META group research note, 2001

[Liang, 12] Liang, X., Li, X., Barua, M., Chen, L., Lu, R., Shen, X. and Luo, H.Y.: Enable Pervasive Healthcare through Continuous Remote Health Monitoring. IEEE Wireless Communications 2012, 1(6), 10-18.

[Luo, 16] Luo, G.: PredicT-ML: a tool for automating machine learning model building with big clinical data. Health Inf Sci Syst 2016, 4-5, doi:10.1186/s13755-016-0018-1.

[Masud, 15] Masud, M., Hossain, M., Alamri, A., Almogren, A., Zakariah, M.: Synchronizing Data through Update Queries in Interoperable E-Health and Technology Enhanced Learning Data Sharing Systems. Journal of Universal Computer Science 2015, 21(11), 1439-1453.

[Moody, 96] Moody, G.B., Mark, R.G.: A Database to Support Development and Evaluation of Intelligent Intensive Care Monitoring. Computers in Cardiology 1996, vol. 23, 657–660.

[Raghupathi, 14] Raghupathi, W. and Raghupathi, V.: Big data analytics in healthcare: promise and potential. Health Information Science and Systems 2014, 2(1), 1-10, doi:10.1186/2047-2501-2-3.

[Ray, 16] Ray, I., Ong, T. C., Ray, I. and Kahn, M. G.: Applying attribute based access control for privacy preserving health data disclosure. Biomedical and Health Informatics(BHI), IEEE-EMBS International Conference 2016, 1-4.

[Raven, 15] Raven, K., Mohan, A., Lopes, R., Solayman, H., Fonseca, P., Jaret, P.: Thinking Healthcare Ahead, 1st ed., Requardt, H. Eds., A Medical Solutions Publication: Erlangen, Germany, 2015, 67-69.

[Sandhu, 96] Sandhu, R. S., Coyne, E. J., Feinstein, H. L. and Youman, C. E.: Role based Access Control Models. IEEE Computer 1996, 2, 38-47.

[Santos, 16] Santos, P., Dennerlein, S., Theiler, D., Cook, J., Treasure-Jones, T., Holley, D., Kerr, M., Attwell, G., Kowald, D., Lex, E.: Going beyond your Personal Learning Network, Using Recommendations and Trust through a Multimedia Question-Answering Service for Decision-support: A Case Study in the Healthcare. Journal of Universal Computer Science 2016, 22(3), 340-359.

[Seo, 15] Seo, D., Lee, M. and Yu, S.: Development of Network Analysis and Visualization System for KEGG Pathways. Symmetry 2015, 7, 1275-1288, doi:10.3390/sym7031275.

[Shameer, 17] Shameer, L., Badgeley, M., Miotto, R., Glicksberg, B., Morgan, J. and Dudley, J.: Translational bioinformatics in the era of real-time biomedical, health care and wellness data streams. Bioinformatics 2017, 18(1), 105–124, doi: 10.1093/bib/bbv118.

[Sun, 13] Sun, J., and Reddy, C.: Big Data Analytics for Healthcare. Tutorial presentation at the SIAM International Conference on Data Mining, Austin, TX, 2013.

[Sunyaev, 10] Sunyaev, A., Chornyi, D., Mauro, C. and Krcmar, H.: Evaluation Framework for Personal Health Records: Microsoft HealthVault vs. Google Health. Proceedings of Hawaii International Conference on System Sciences, Hawaii, USA, Jane 05-08, IEEE Computer Society: Washington DC, USA, 2010, 1-10.

[Wang, 04] Wang, M., Lau, C., Matsen, F. and Kim, Y.: Personal Health Information Management System and its Application in Referral Management. IEEE Transactions on information technology in biomedicine 2004, 8(3), 287-297.

[Wang, 16] Wang, L., Dong, X., Zhang, X., Guo, F., Wang, Y. and Gong, W.: A Logistic Based Mathematical Model to Optimize Duplicate Elimination Ratio in Content Defined Chunking Based Big Data Storage System. Symmetry 2016, 8, 69, doi:10.3390/sym8070069.

[Zhang, 04] Zhang, G. and Parashar, M.: Context-aware dynamic access control for pervasive applications. In Proceedings of the Communication Networks and Distributed Systems Modeling and Simulation Conference 2004, 1-6.

[Zheng, 14] Zheng, Y., Ding, X., Poon, C., Lo, B., Zhang, H., Zhou, X., Yang, G., Zhao, N. and Zhang, Y.: Unobtrusive Sensing and Wearable Devices for Health Informatics. IEEE Transactions on Biomedical Engineering 2014, 61(5), 1538-1554.