Privacy Protection Issues for Healthcare Wellness Clouds

Tyrone Grandison[*], Pei-yun S Hsueh[*], Liangzhao Zeng[*], Henry Chang[+], Yi-Hui Chen[#], Ci-Wei Lan[+], Howard Pai[+], Li-Feng Tseng[+]

[*]IBM Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598, USA

[#]Dept. of Information Science and Applications, Asia University, No. 500, Lioufeng Road, Wufeng, Taichung County 41354, Taiwan

[+] Taiwan Collaboratory, IBM Taiwan Corporation, P.O. Box 4F, No. 7, SongRen Road, 1073, Taipei, Taiwan, R.O.C.

ABSTRACT

Healthcare is ubiquitous in every business organization. Whether as the primary focus of the business or as a function of the well-being of a firm's employees, health issues play a dominant role in commerce. This recognition and the demonstrated benefits of a healthy contributor or worker have promoted a rejuvenated emphasis on wellness. In order to garner the benefits of cloud computing and foster improved employee health, the Taiwan Collaboratory is developing a first instance of a Wellness Cloud, which is an integrated, interconnected and intelligent well-being platform. As the data held in this cloud is potentially very sensitive, the protection of this data is of utmost importance. In this chapter, we present issues and solutions for protecting user data while enabling the data to be usefully processed and for value to be derived, by using advanced technology and by harnessing the cumulative knowledge or wisdom of the collective of users.

INTRODUCTION

Though there is no universally accepted definition of wellness (Baranowski, 1981; Savolaine & Granello, 2002), it is generally acknowledged that the abstract concept of wellness centers around *the active process of becoming aware of and making choices toward a more successful existence* (Mackey, 2000; Corbin & Pangrazi, 2001). Physical wellness involves the collection, analysis and presentation of actionable personal information over time in order to help patients prevent illness, positively manage current conditions, and make healthier choices.

There is a rich history of forwarding-thinking governments and enterprises that offer wellness incentive programs (Goetzel et.al., 1994; Maes, 1998; Ozminkowski et.al., 2002; Loong, 2009; Nakamura, 2010), such as smoking cessation rebates, weight management and fitness goal rewards. Unfortunately, the

current set of wellness initiatives and tools are generally siloed solutions that are refreshed annually and are not integrated with other wellness and health management systems.

BACKGROUND

In late December 2009, the Taiwanese government embarked on a project to leverage novel technologies in addressing the wellness needs and desires of the people of Taiwan (Zane, 2009). Partnering with IBM, they established a collaboratory, which is a (virtual) laboratory where IBM researchers worldwide co-locate with local universities, government, or commercial partners to share skills, assets, and resources to achieve a common research goal (Nystedt, 2009).

The Taiwan Collaboratory utilizes cloud computing systems, remote monitoring technologies and advanced user interface techniques and methodologies to ingest and integrate large volumes of data on multiple aspects of the (current) condition of a citizen. This data is combined with the citizen's historical data to provide invaluable feedback to them on their continued progress towards their pre-specified goal or to help them recognize when they are on a path to unfavorable outcomes. The data is also securely leveraged with the data of others to perform advanced analytics in order to detect trends and insight that would have otherwise been undiscoverable.

As wellness is a sub-discipline of the broader healthcare domain that has only been examined and implemented as manual, human-intensive efforts, there are no comparable wellness software solutions with similar goals to the collaboratory. However, the closest Personal Health Record (PHR) systems that may be thought of as offering similar functionality are Microsoft HealthVault (http://www.healthvault.com), Google Health (http://www.google.com/health/) and Personal Care Connect (Blount et.al., 2007). Both Google Health and Microsoft HealthVault are healthcare information portals that allow patients to 1) consolidate their information from disparate data sources, 2) set personal health goals, 3) track their progress, 4) share their health information, and 5) enable the companies' partners to access patient data and provide services. Both systems do not support real-time monitoring of a patient's wellness state and the internal analysis and processing of incoming data for positive and negative trends. Additionally, the security and privacy safeguards utilized throughout their ecosystems are steeped in obscurity and supported by a trust model that is rooted in relying on the "goodness" of each company's brand. Personal Care Connect (PCC) is a standards-based, open solution that was developed by IBM to facilitate the remote monitoring of patients in order to provide timely access to a patient's health status. Though, PCC addresses the real-time monitoring deficiency of the previous two solutions, it still suffers from the lack of native advanced analytics and the opacity in privacy and security.

It should also be noted that all these systems must comply with legislative policy rules that stipulate privacy and security mandates. The Health Insurance Portability and Accountability Act (HIPAA, 1996) is the regulatory foundation in the United States. It establishes safeguards to protect the privacy of individually identifiable protected personal health information (PHI), sets limits and conditions on the uses and disclosures that may be made on PHI with and without patient authorization, and gives patients rights over their PHI. The Health Information Technology for Economic and Clinical Health Act (HITECH, 2009) and the healthcare-specific amendments in the American Reinvestment and Recovery Act (ARRA) has enhanced HIPAA over the last few years.

The construction of a Wellness Cloud is a bold instance of the building an integrated, interconnected and intelligent wellness care system that is focused on helping people with their personal wellness goals, the discourse in this chapter is meant to be instructive (not exhaustive or definitive) in the development and or use of similarly purposed systems. More specifically, the general focus of this work is to provide a blueprint for others on the privacy issues involved and the protection mechanisms that can be used to address these concerns.

WELLNESS CLOUD ISSUES

This undertaking requires an ecosystem of many stakeholders, from different backgrounds, with varying motivations, but with the same common purpose – to collaborate and use wellness information to make people better informed, and to improve and expand the services that can be offered by the system (or its constituent stakeholders). Some of the stakeholders in the Wellness Cloud include medical device manufacturers, fitness outlets, healthcare providers and government administrators. The insight harnessed from the amalgamation of the data acquired or generated by each partner is the most important value delivered to the cloud's end-users. It is this value and the power of data compounding that are the driving force behind partners' decision to participate in this cloud.

However, as Personally Identifiable Information (PII) will be generated from each party, mechanisms are utilized to ensure the security and privacy of the information as it is communicated from the partner to the cloud and vice versa. Once, PHI is ingested into the Wellness Cloud, the information is transformed in order to be compliant with international standards for representing and storing healthcare data. Then, techniques are employed to ensure that data integration, data processing and analytics are performed in a privacy-preserving manner. When information is requested by a citizen or other stakeholders, policy-driven techniques for protecting the rendered information are used to ensure that there is a low probability of inadvertent information leakage.

In the chapter, we discuss all the areas of concern with regards to the privacy of information in the Wellness Cloud, and we provide an analysis of the best practices for protecting against these concerns; stating the approach employed by the Taiwan Collaboratory. As this effort is, at the core, a dynamic, virtual business organization constructed from multiple businesses, we envision that the issues and techniques presented will be applicable not only to this scenario, but to many other virtual business collaborations. As with all other systems, we start with a description of the core architectural components.

THE WELNESS CLOUD ARCHITECTURE

Enabling wellness management on a cloud system requires that particular (user) expectations must be axioms of the computing platform. Some of the more interesting expectations or requirements include mechanisms for personalization, native handling and management of events, support for dynamicity, and the ability to be scalable in a number of dimensions. These imperatives preclude the use of the current set of publicly available cloud platforms, such as the Amazon Web Service (Lerner, 2006), Eucalyptus (Nurmi, 2008) and Force.com (Tibken, 2010) for wellness management.

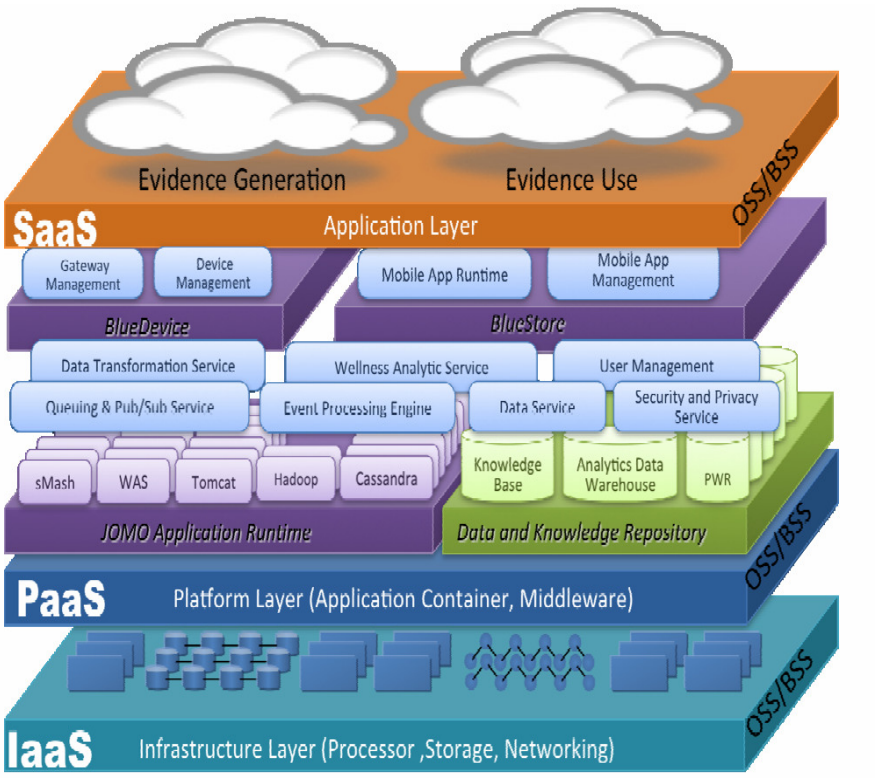

Figure 1. GreenOlive Cloud Platform for Wellness Management

In order to tackle these challenges, the team developed the *GreenOlive* cloud platform (Figure 1), which consists of three layers:

(i) Infrastructure Layer – this layer provides Infrastructure-as-a-Service (IaaS) for platform services. The infrastructure resource includes processing, storage and networking that are required by provisioning the platform services.

(ii) Platform Layer – this layer consists of an application runtime platform (codenamed JOMO), a collection of data & knowledge repositories, a collection of services that run on top of JOMO, and data & knowledge repositories. The services include queuing and publish-subscribe (hereafter referred to as pub-sub) services that provide communication channels among services; security and privacy services that ensure that user privacy is protected; data transformation services that transfer different data/event sources to standard formation; event processing services that manage events in real-time; data services that provide access to Application Programming Interfaces (APIs) for data and knowledge repository; user services that manage user information; and wellness analytic services that generate new guidelines or new insights of existing guidelines for wellness management. It should be noted that each of the above services provide a set of open APIs that allow partners to develop new services or applications. On top of these services, two management services (i.e. BlueDevice and BlueStore) are created, wherein BlueDevice provides gateway and device management and BlueStore provides mobile application runtime and management services.

(iii) Application Layer – this layer adopts and implements the Software-as-a-Service (SaaS) paradigm to provision applications. These applications are developed using the open APIs on the platform

layer. Further, the platform provides application composition or mashup mechanisms that enable multiple applications to form an ecosystem. This then enables the development of the novel offering of an Ecosystem-as-a-Service (EaaS). In particular, our solution focuses on two categories of application, i.e. evidence generation and evidence use. Evidence generation aims to generate new clinical evidence, guidelines and insights based on collected data in the cloud, while evidence use is concerned with the delivery of clinical knowledge to users, based on clinical context information.

The interested reader can peruse Hsueh, et.al. (2010) and Zeng et.al. (2010) for a more detailed description of the current and future states of the platform. Using our architectural discussion as a base, we proceed with a discussion of the protection issues involved in data collection, as data is transported and distributed within the cloud and when it is used.

PRIVACY PROTECTION IN DATA COLLECTION

Figure 2 shows the typical ingestion model for cloud computing systems. In the contemporary ingestion model, it is assumed that the information being sent to the cloud only needs to be safeguarded once it enters the cloud. This assumption cannot be made for wellness management clouds, because wellness data typically have higher levels of sensitivity and significantly higher risk profiles.

Figure 2. Standard Data Collection in a Cloud

There are a number of other privacy factors that highlight that the conventional cloud data collection paradigm is insufficient. These factors include: 1) the users' privacy expectations, 2) the legislative mandates to protect PII and, 3) the latent objective to foster and ensure openness, accountability and trust in the user population in order to ensure future system use. The compromise of any of the factors has the potential to negatively impact platform sustainability. With these issues in mind, design decisions must be made to ensure that the data is protected from the point of collection to the point of insertion into the cloud (Figure 3).

Security controls must be in place when data is captured at the human-machine interaction point (whether it is a medical device, custom-made web portal, mobile platform or conventional computer), when data

travels from the interaction point to the cloud ingestion point via the communication channel, and when it is about to be ingested into the cloud. The data falls into one of the following categories: demographic, body measurement, past history, current activity and social & psychological; and the decision to protect each piece of data is dependent on its sensitivity, which is determined a priori and periodically re-evaluated. As shown in Figure 3, the Security and Privacy (S&P) Transformer is the secure gateway for all data to be ingested into the Wellness Cloud.

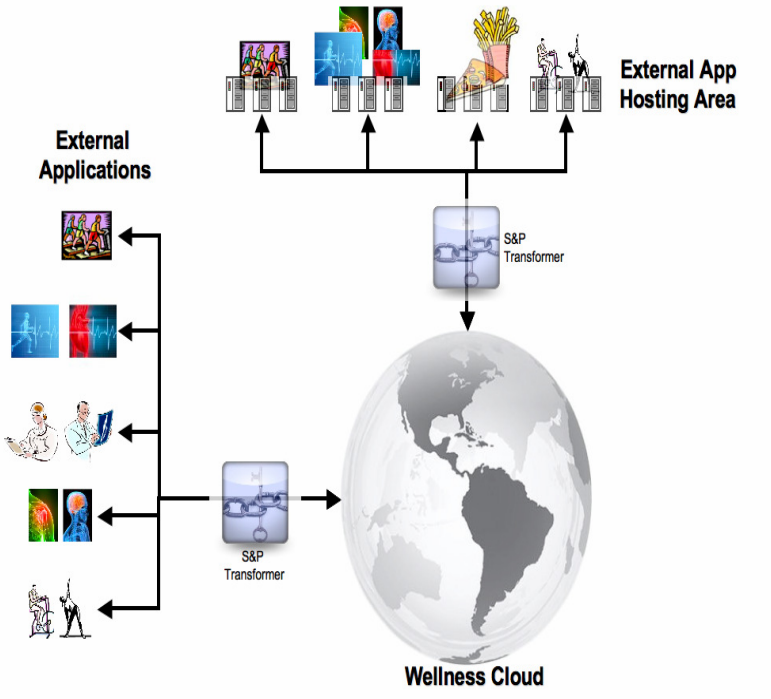

Figure 3. Wellness Cloud Ingestion Model

Generally, for end-to-end preservation of privacy in a networked environment (Beresford & Stajano, 2003; Chan & Perrig, 2003; Mokbel, 2007; Langheinrich, 2009), controls may be required on the identity (i.e., of the sender or recipient), content (i.e. the message being sent) or the context (i.e. details, such as message or identity metadata, that can reveal or lead to inference on private information).

In order to safeguard identity, protection techniques, such as Mixes (Chaum, 1981) and Onion Routing (Goldschlag, 1999), have been proposed. A Mix is a mediator between service consumers and service providers that performs cryptographic transformations on incoming messages, and then forwards the messages to the relevant party. Onion routing refers to the idea of using a set of onion routers, where each router unwraps a layer of encryption around a message that was repeatedly encoded to reveal routing instructions. This prevents intermediary nodes from knowing the source, destination or message. In the context of a Wellness Cloud, identity is needed for internal processing and routing of updates to the appropriate agents. For example, it is often useful to have the results of your cardio-vascular session sent to your personal wellness record or to your doctor. Thus, these techniques are not ideally suited for this environment.

For content protection, using public key cryptography with strong encryption keys is regarded as the best practice. Current mechanisms used in this space are SSL (Secure Socket Layer) and HTTPS (Hypertext

Transfer Protocol Secure). However, Bissias et.al. (2005) have demonstrated that it is possible to successfully perform traffic analysis attacks on encrypted HTTP streams. Thus, there is further research to be done on this topic.

In terms of context protection, location data is the contextual attribute that has been the most extensively studied. Gruteser et.al. (2003) promote the application of a distributed anonymity algorithm before access is granted to a service provider. Ozturk et.al. (2004) propose augmented routing protocols to protect the location of the source during sensor transmissions. Gedik and Ling (2005) propose an information sharing framework that allows individual nodes to specify how large a group they wish to hide in and then generalizes location based on the group size. In practice, contextual attributes other than location still require investigation. For the Wellness Cloud effort, location may be relevant for analytic services. Thus, that specific contextual attribute is included as is.

For the Taiwan Collaboratory, each device manufacturer employs their own techniques to secure the sensitive data, i.e. identifiers and quasi-identifiers, on the device. The partner then engages a secure channel to the gateway of the Wellness Cloud (WC) – this is the S&P Transformer (SPT) or the Security and Privacy Transformation Unit previously mentioned. Currently, this secure channel is a SSL (Secure Sockets Layer) session and the SPT transforms incoming data into its anonymized form, using a set of rules gathered from analysis of legal requirements and customer requests.

It should be noted that the physical location of the SPT is a design decision that impacts the overall security and privacy of the system. A SPT host that is a trusted authority with a dedicated and direct line to the WC would be the optimal configuration for risk and liability reduction. A SPT host that is co-located with the WC would be the most system-efficient and would reduce the likelihood of data exposure or disclosure.

In order to ensure that each device manufacturer is identifiable within the cloud and that only de-identified information is held and processed by the WC, the SPT utilizes an algorithm that randomly selects a transformation algorithm (TA) for the data stream based on a set of identifier modification options. The system maintains a pool of TAs that perform tasks that range from simple transpositions (e.g. switching male to female for gender) to more complex functionality (e.g. converting diagnosis and condition information into their abstract form via ICD10 – International Classification of Diseases, 10th Edition). Based on the device id, the patient identifier and the incoming data, TAs are chosen and applied to both identifiable information and possibly sensitive information. A hash function is selected to create pseudonyms for identifiable information and TAs are chosen to transform possible sensitive information. The metadata for the mappings are stored by the SPT and are utilized when information needs to be sent from the cloud to service consumers in its raw form. Though the pseudonym life cycle management is simpler than contemporary approaches (Lysyanskaya, 2000), its simplicity is well suited for a system with a large number of users and demanding response time requirements. The trust model behind the pseudonymisation approach is beyond the scope of this chapter.

After ingestion, there are further controls in place within the cloud to further reduce the security and privacy risk.

PRIVACY PROTECTION WITHIN THE CLOUD

There are two key issues to be addressed within the cloud. The first issue occurs when data is being transported and distributed to the difference services inside the cloud. As these data may contain residual private information (i.e. groups of seemingly innocuous data items that can be collated and used to uniquely identify people), they should only be distributed to those services that provide services for the user. The second issue occurs when there is information that needs to be sent to users, where this information may also contain (residual) private information that should only be delivered to a specific person and related family members. Both of these concerns are addressed in the literature by using rule-based enforcement mechanisms either at the database layer (Agrawal et.al., 2002), at the application layer (Pearson et.al., 2009; Pearson, 2009; Wang, 2009) or the network layer (del Alamo et.al, 2010). The Wellness Cloud currently utilizes the filtering and transformation rule approach to policy enforcement at the database level.

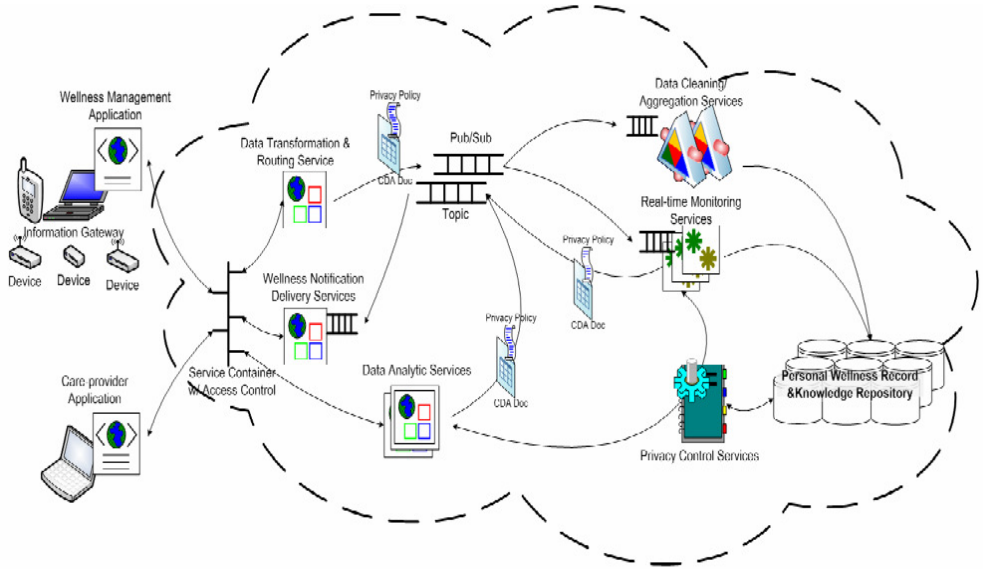

Figure 4. Data Transportation and Distribution

Figure 4 illustrates some data transportation and distribution scenarios within the cloud. The scenarios include two kinds of applications: (i) Wellness Management Applications, and (ii) Care Provider Applications. Wellness Management Applications collect information and tend to be interfaces developed by device manufacturers or wellness program developers. Care Provider Applications enable medical practitioners to help patients achieve their wellness goals and manage their general health.

When the data are collected and sent to the cloud infrastructure, we assume all data will be translated to the CDA (Clinical Document Architecture) format via the *Data Transformation and Routing Services* component. In the WC, multiple CDA documents, when combined, may unlock different levels of private information at a more complete or aggregate level. For example, one CDA document may contain care data and clinical environment data (in various transformed states) that may be innocuous in their own streams, but may become more sensitive as each person's wellness record grows over time; due to the input of system partners. Another document may only contain de-identified information.

When the system distributes these CDA documents inside the cloud infrastructure, privacy polices are adopted to describe who can receive CDA documents. For example, if a CDA document contains all the private information of a user, it should only be delivered to her or his own primary care physician or

applications and services operated by her or his primary care physician. While a CDA document that only contains completely anonymous information may be distributed to any care providers who are interested in clinical data encoded in CDA documents and who agree to comply with the cloud's data use agreements.

When notification or feedback information (encoded as CDA documents) is created by wellness services (e.g., real-time monitoring services), these CDA documents also contain private information. Therefore, privacy control in distributing these CDA documents is also required. For the Wellness Cloud, fine-grained, data-level protection technologies (Bird et.al., 2005; Grandison et.al., 2007) are used in privacy policy management in order to ensure disclosure compliance. For example, if the CDA documents contain users' personal information, the documents should be delivered only to users, their family member and their primary care physicians. The policy management tools allow the patient to state the exact set of people they want to have access rights to their data.

Now we discuss the privacy policy enforcement in information distribution. In our solution, we adopt a pub-sub mechanism to distribute the CDA document. In a typical pub-sub system, the information providers publish messages that information consumers subscribe to by registering subscriptions. There are three kinds of pub-sub systems, namely topic-based, attribute-based and content-based. In topic-based pub-sub, all the messages published to a topic are delivered to message consumers who subscribe the topic. In attribute-based pub-sub systems, each message is associated with a collection of attributes, while a subscription is a Boolean predicate on attributes. When a message is published, if the consumer's Boolean predicate is evaluated to true, the message is delivered to the subscriber. Different from attribute-based pub-sub, in content-based systems, the Boolean predicate covers the entire content of the whole message. In our system, the above three pub-sub systems are adopted for different kinds of information distribution among users, services and service providers. However, existing pub-sub systems do not support privacy policy enforcement. In our system, when the pub-sub system is matching the message and subscriptions, it will first use privacy policy filters on the subscriptions. For example, a CDA document is associated with a policy that states that only user's primary care physician can access it. Therefore, when matching the CDA document with a subscription, the system retrieves the user's primary care physician list and uses the list to filter the subscription before evaluating the subscription predicates.

PRIVACY PRESERVATION IN WELLNESS ANALYTICS

One important motivation of the Wellness Cloud design in the Taiwan Collaboratory is to support the dynamic formation of wellness ecosystems. The expectation is that this will facilitate the provisioning of personalized services and satisfy the long tail of wellness care needs in an economically viable way. Typically, wellness ecosystems are composed of multiple service providers, each responsible for one part of the collaborative care solution. The Wellness Cloud architecture is therefore designed to enable streams of patient data to be transported and distributed to different service providers in a secure way. Nevertheless, the data obtained from various providers proves to be challenging for the stakeholders to comprehend. Consequently, wellness analytics services have been proposed to help transform the collected patient data into actionable knowledge for stakeholders.

Three common wellness analytic services have been identified and are supported. The first helps determine who is in need and what major risk factors should be attended for further intervention. For example, a case manager needs to be alerted if any of the residents in a long-term care institute are developing an increased risk for chronic conditions. The second enables providers to learn from a user's personal wellness history so as to provide the user with personal recommendations of intervention plans. For example, registered dietitians in an out-patient nutrition service center, when given information about their patients' nutrition compliance history, could better select education materials and adjust meal plans for the patients. The third leverages wellness status determination analysis services in an online fashion to monitor the change in a user's wellness status and to introduce the selected intervention plan at the point of change. For example, the recognition of a significantly lowered heart rate and low blood sugar levels by the WC may trigger a call to the closest ambulatory unit, if the patient has a history of strokes.

Even though all these services use only datasets that conform to legislation, such as the Health Insurance Portability and Accountability Act (HIPAA), there still exist risks of privacy intrusion (Liu & Terzi, 2009; Maximilien et.al., 2009; Becker & Chen, 2010). Many prior cases have shown that important personal information can still be recovered with carefully crafted queries from de-identified records during analysis. For one, Sweeney (2002) has identified the medical records of the former governor of Massachusetts by cross-referencing the voter registration lists and the anonymous National Association of Health Data Organizations (NAHDO) hospital records. To protect dangerous information leakage, previous research has proposed various auditing and data perturbation methods. Some propose auditing tools to restrict queries and avoid malicious attacks. Others mask or introduce noise to randomize the input dataset (Kim and Winkler, 1997; Lindell and Pinkas, 2002; Wilson and Rosen, 2003; Loukides and Shao, 2006), or return a noisy version of output analysis (Dinur and Nissim, 2003; Chen et al., 2007). While these methods all have their merits, they are designed with the assumptions that service providers own the data they ingest and protection mechanisms are skewed in the favor of providers (Grandison, 2010). In the context of cloud-based wellness ecosystems, usually more than one data source is used for analysis. It is therefore impractical to put any single service provider in charge of overseeing the privacy issue.

The challenges and opportunities incurred by the multi-provider environment on the privacy protection mechanism in a Wellness Cloud are two-fold. On the one hand, although it is difficult for single service providers to measure the level of privacy risk associated with a particular individual, the auditing mechanism can be extended to operate in the cloud and actively avoid choosing the records belonging to those who are already at a high privacy risk, i.e. easily identifiable individuals, for analysis. On the other hand, beyond legislative regulations, the data owners, i.e. patients, may all have different privacy requirements for their own personal information. There is currently no easy way for the data owners to specify their requirements. In the remainder of this section, we will continue the discussion of extending the cloud-based auditing mechanism to actively monitor the privacy risk and select the right mitigation strategy according to patient's specifications.

Privacy-preserving Active Sampling for Risk Modeling

Most wellness analytic services originate from the need to infer incoming users' wellness status, i.e., the propensity to various conditions. Our prior work has focused on designing a large-scale distributed

infrastructure that can monitor the wellness information data streams of a large population and learn how to discriminate the status of incoming data streams from previous records (Zeng at al., 2010). To satisfy the long tail of demand with personalized services, feature selection and sampling approaches are needed to scan across the available databases and select subsets of data to develop models for characterizing wellness status of the target individual (Hsueh et al., 2010). The developed models based on risk grouping are then used to single out the major risk factors associated with a target individual and provide personalized recommendations of suitable follow-up intervention plans.

Because the personalized wellness analytics services are designed to use as little data as possible for economic reasons, we propose the utilization of an active sampling framework, which aims to find the smallest subset of data that is sufficiently representative of the risk group without degrading the performance on wellness status estimation. Previous proposals include using a filter approach to identify a subset of data that exhibit the strongest global utility for describing the target wellness status. However, without considering privacy issues, many of the personal wellness attributes of data owners in a risk group can be automatically inferred, e.g., by majority voting. With the privacy issue in consideration, the active sampling framework is then recast as follows:

> **Input:** Data records in the same risk group □*(f1,…fn, S)*; the privacy requirement θ (which is represented as the maximum allowed number of inferable attributes);
>
> **Output:** the subset of records that maximize the wellness status association without compromising the privacy requirement.

The development of such an active sampling framework requires the auditing mechanism in the back end (usually run by the cloud operator) to record the personal wellness attributes associated with a particular data owner. The privacy risk is estimated by the number of inferable attributes with the selected data records of the same risk group.

For the online version of wellness analytic services, the auditing mechanism is extended to monitor the changes in privacy risks, using a sliding window approach. The privacy risk at time *t-1* and *t* are profiled as y_{t-1} and y_t. The wellness attributes (which, in combination, may represent different levels of privacy risks) are represented as F_{t-1} and F_t, and the risk factors of privacy level change (which describe events associated with changes in privacy risk levels) as $P_c(F_{t-1})$ and $P_c(F_t)$. As shown in Formula (1), the goal is to search for a privacy risk level pair that simultaneously maximize $P_A(F_{t-1}, y_{t-1})$ and $P_A(F_t, y_t)$, i.e., the likelihood of the estimated privacy risk level at the two time points, and minimize the penalty of risk level change from time *t-1* to *t*, i.e., the likelihood of risk change at the two time points $(w_2 P_c(F_{t-1}) + w_3 P_c(F_t))$ and the imposed penalty on risk change, ϕ. The weightings of the terms in the scoring function are learned from privacy intrusion history. If there is a privacy risk level change, the system will raise alerts to the service provider for mitigation.

$$g(< y_{t-1}, y_t >) = w_1 P_A(F_{t-1}, y_{t-1}) + w_1 P_A(F_t, y_t) + \phi(y_{t-1} \neq y_t)(w_2 P_c(F_{t-1}) + w_3 P_c(F_t))$$

Formula (1)

Having measured the amount of information leakage associated with a particular person, it then comes down to the question of how to identify the right risk mitigation strategy for the active sampling framework. This is a problem that is difficult for any single data provider to handle on their own, but can be rendered more easily on Wellness Clouds with the aid of a personal privacy requirement handler.

Decision Support Framework for Personal Privacy Requirement Handler

Selecting the right privacy risk mitigation strategy for an individual data owner requires understanding what composes a privacy threat to the individual. Different countries have different regulations concerning individual privacy rights. Moreover, data owners under similar circumstances may perceive privacy threats differently; even for the same wellness attribute, sometimes they would weigh the importance of information leakage differently. For example, a recently diagnosed diabetes patient may prefer more protection against the leakage of related information than long-time patients. Therefore, a personal privacy requirement handler should allow data owners to specify their privacy requirements. Default country-specific or institute-specific requirement templates can be provided by the privacy-responsible authorities.

The selection of risk mitigation strategy can then be characterized as an inference problem. Given the alternative action plans (*AP*) following different strategies, each represented as a set of plan features f_i, the goal is to rank the action plans according to how well a plan serves to address the specified privacy threat of the data owner at risk. The importance of ap_t (*t*th action plan in consideration) on u_a (*a*th user) is measured with the multiple-attribute value function:

$$imp(u_a, ap_t) = \sum_{f_i \in ap_t} w_{f_i}' imp(u_a, f_i) \qquad w_{f_i}' = w_{f_i} \Big/ \sum_{i=1}^{n} w_{f_i}$$

Formula (2)

where f_i is the *i*th privacy risk factor considered in ap_t, and w_{f_i}' is the owner-specified weight on f_i.

If the privacy threat is not specified by the data owner, the system then applies a collaborative filtering approach to rank action plans, using the requirements of data owners in the same risk group, *N*. The importance of a factor is determined by the weighted average of importance ratings, adjusted by the similarity between the owners.

$$imp\ (u_a, f_i) = \frac{\sum_{n \in N} sim\ (u_a, u_n) imp\ (u_n, f_i)}{\sum_{n \in N} sim\ (u_a, u_n)}$$

Formula (3)

The analytics services can start offering privacy handling with the auditing mechanism, privacy-preserving active sampling framework, and personal privacy requirement handler in development. However, there are still questions remaining. For one, we only implement two threat models: owner-specified and collaborative filtering-based attribute weighting. Taking a step back, we need to understand how to model the privacy threats. Different scenarios inherit different tradeoffs between privacy and analysis requirements. Yet the sampling approach may come with different granularity: e.g., spatial, temporal, and identity. How do we estimate temporal changes effectively and efficiently without having

to compromise the privacy threshold set by each individual in the sample set? Also, recent research shows that unsupervised clustering approaches can effectively group utility functions of individuals into a prototypical function according to how close they are to each other in a multi-dimensional space. The unsupervised approach can further assist the development of the online version of the privacy protection mechanism. In addition, owing to the nature of wellness services invoked by vital sign monitoring devices (e.g., for analyzing abnormal glucose patterns after meals), the system needs to support flexible and scalable service provisioning. This includes the support of high-throughput privacy-preserving risk grouping and risk mitigation through the cloud-based load balancing mechanism.

The privacy risk mitigation mechanism proposed here has impacts on wellness education and monitoring. With the better privacy control mechanism in place, the wellness analytics service sets the stage for a new generation of personal wellness decision support systems, which aim to reinstate individuals' self-assessment capabilities and a better sense of control over their own wellness management process.

IMPLICATIONS AND FUTURE WORK

For the average business organization that is deploying a cloud, care must be taken in the design and implementation of the privacy and security controls at data collection, during transit and while it is being held in the cloud. Often, this may require a collaboration with the partners that produce the devices or the platform from which data is to be ingested and an agreement with the affiliate service consumers that build services on top of the cloud.

The strategic research initiatives involve 1) infusing specialized cryptographic schemes in the cloud data ingestion and transformation processes, 2) building more intuitive and friendly interfaces for policy acquisition and management, 3) innovating sophisticated risk management algorithms and technologies, and 4) moving toward a software-hardware hybrid system where specialized hardware components are included to handle (computationally intense) CDA processing and execute cryptographic methods.

Tactically, it is expected that there will be lessons learned from the current deployments as the system continues to be used and users make requests to improve their experience and interaction.

CONCLUSION

In this chapter, we introduced the Wellness Cloud – an integrated, interconnected and intelligent healthcare well-being system – that is developed to help citizens achieve their wellness goals. We presented the issues around privacy protection on wellness devices, while data is being transmitted from device to cloud, when it is being processed within the cloud and while it is being used by analytic services. We also presented the approaches taken and highlight the future direction of the effort.

It is our hope that this articulation will serve as 1) a spark for discussion, and 2) a template for entities who either want to develop ecosystem components or similarly purposed systems.

REFERENCES

Agrawal, R., Kiernan, J., Srikant R., & Xu, Y. (2002). Hippocratic Databases. Proceedings of the 28th Int'l Conf. on Very Large Databases, Hong Kong, China.

Agrawal, R., & Srikant, R. (2000). Privacy-Preserving Data Mining. Proceedings of the ACM SIGMOD Conference on Management of Data, Dallas, Texas, USA.

Baranowski, T. (1981). Toward the definition of concepts of health and disease, wellness and illness. Journal of Health Values, *5*(6), 246-256.

Becker, J. & Chen, H. (2010). Measuring Privacy Risk in Online Social Networks. Proceedings of the Web 2.0 Security and Privacy (W2SP).

Beresford, A.R., & Stajano, F. (2003). Location Privacy in Pervasive Computing. IEEE Pervasive Computing, *2*(1), 46- 55.

Blount, M., Batra, V. M., Capella, A. N., Ebling, M. R., Jerome, W. F., Martin, S. M., Nidd, M., Niemi, M. R., & Wright, S. P. (2007). Remote health-care monitoring using Personal Care Connect. IBM Systems Journal, *46*(1), 95-113.

Bird, P., Grandison, T., Kiernan, J., Logan, S., & Rjaibi, W. (2005). Extending Relational Database Systems to Automatically Enforce Privacy Policies. Proceedings of the 21st International Conference on Data Engineering (ICDE), Tokyo, Japan.

Bissias, G.D., Liberatore, M. & Levine, N.B. (2005). Privacy Vulnerabilities in Encrypted HTTP Streams. Proceedings of the Privacy Enhancing Technologies Workshop, Dubrovnik, Croatia.

Chan, H., & Perrig, A. (2003). Security and Privacy in Sensor Networks. IEEE Computer, *36*(10), 103-105.

Chaum, D. (1981). Untraceable Electronic Mail, Return Addresses and Digital Pseudonyms. Communications of the ACM, *24*(2), 84–88.

Chen, B., Lefevre, K., & Ramakrishnan, R. (2007). Privacy skyline: Privacy with multidimensional adversarial knowledge. Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB), University of Vienna, Austria.

Corbin, C.B., Pangrazi, R.P. (2001). Towards a Uniform Definition of Wellness: A Commentary. President's Council on Physical Fitness and Sports. Series 3, No. 15. Retrieved April 4, 2011 from http://eric.ed.gov/PDFS/ED470691.pdf

Dasu, T., Krishnan, S., & Venkatasubramanian, S., Yi, K. (2006). An Information-Theoretic Approach to Detecting Changes in Multi-Dimensional Data Streams.

del Alamo, J.M., Monjas, M.A., Yelmo, J.C., San Miguel, B., Trapero, R., & Fernandez, A.M. (2010). Self-service Privacy: User-Centric Privacy for Network-Centric Identity. Proceedings of the IFIP Advances in Information and Communication Technology, *321*, 17-31.

Dinur I., & Nissim, K. (2003). Revealing Information while Preserving Privacy. Proceedings of the Twenty-Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS), San Diego, California.

Gedik, B., Ling, L. (2005). Location Privacy in Mobile Systems: A Personalized Anonymization Model. Proceedings of the 25th IEEE International Conference on Distributed Computing Systems (pp. 620-629).

Goetzel, R., Sepulveda, M., Knight, K., Eisen, M., Wade, S., Wong, J., & Fielding, J. (1994). Association of IBM's "A Plan for Life" Health Promotion Program with Changes in Employees' Health Risk Status. Journal of Occupational Medicine, *36*(9), 1005-1009.

Goldschlag, D., Reed, M., & Syverson, P. (1999). Onion Routing. Communications of the ACM, *42*(2), 39-41.

Grandison, T. (2010). Patient-Centric Privacy: Envisioning Collaboration Between Payers, Providers & Patients With The Patient At The Core. Proceedings of the 6th IEEE International Conference of Collaborative Computing: Networking, Applications, and Worksharing (CollaborateCom), Chicago, Illinois.

Grandison, T., Johnson, C.. & Kiernan, J. (2007). Hippocratic Databases: Current Capabilities and Future Trends. Handbook of Database Security: Applications and Trends. Editors: Sushil Jajodia and Michael Gertz.

Gruteser, M., Schelle, G., Jain, A., Han, R., & Grunwald, D. (2003). Privacy-Aware Location Sensor Networks. Proceedings Of The 9th Conference On Hot Topics In Operating Systems (HOTOS), Berkeley, CA, USA.

Health Insurance Portability and Accountability Act (HIPAA). (1996). Retrieved April 4, 2011 from http://www.intellimark-it.com/privacysecurity/hipaa.asp.

Health Information Technology for Economic and Clinical Health Act (HITECH). (2009). Retrieved April 4, 2011 from http://www.hhs.gov/ocr/privacy/hipaa/administrative/enforcementrule/hitechenforcementifr.html.

Hsueh, P., Lin, R., Hsiao, J., Zeng, L., Ramakrishnan, S. & Chang, H.. (2010). Cloud-based Platform for Personalization in a Wellness Management Ecosystem: Why, What, and How. Proceedings of the 6th IEEE International Conference of Collaborative Computing: Networking, Applications, and Worksharing (CollaborateCom), Chicago, Illinois.

Kim, J. & Winkler, W. (1997). Masking Microdata Files. Technical Report of Bureau of the Census. Retrieved April 4, 2011 from http://www.census.gov/srd/papers/pdf/rr97-3.pdf

Langheinrich, M. (2009). A Survey of RFID Privacy Approaches. Journal of Personal Ubiquitous Computing, *13*(6), 413-421.

Lerner, R.M. (2006). Amazon Web Services. Linux Journal, *143*, 20-24.

Lindell, Y. & Pinkas, B. (2000). Privacy Preserving Data Mining. Proceedings of the Advances in Cryptology (pp. 20-24), LNCS 1880, Springer-Verlag.

Liu, K. & Terzi, E. (2009). A Framework for Computing the Privacy Scores of Users in Online Social Networks. Proceedings of the IEEE International Conference on Data Mining (ICDM). Miami, Florida. USA.

Loong, L.H. (2009). Preparing for an aging population - The Singapore experience. The Journal: AARP International (Winter), 12-17. Retrieved April 4, 2011 from http://www.nus.edu.sg/nec/InnoAge/documents/AARPjournalwinter09_PMLee.pdf

Loukides G. & Shao, J. (2008). Data Utility and Privacy Protection Trade-off in k-Anonymisation. Proceedings of the First International EDBT Workshop on Privacy and Anonymity in the Information Society (PAIS), Nantes, France.

Lysyanskaya, A., Rivest, R. L., Sahai, A., & Wolf, S. (2000). Pseudonym Systems. Selected Areas in Cryptography - Lecture Notes in Computer Science, 2000, *1758*, 184-199.

Nakamura, D. (2010). Fat in Japan? You're breaking the law. Retrieved April 4, 2011 from http://www.globalpost.com/dispatch/japan/091109/fat-japan-youre-breaking-the-law

Nystedt, D. (2009).Taiwan to Host IBM's First Joint Healthcare IT Research Unit. IDG News Article. Retrieved April 4, 2011 from http://www.pcworld.com/article/185193/taiwan_to_host_ibms_first_joint_healthcare_it_research_unit.html

Mackey, S. (2000). Towards a definition of wellness. Journal of Holistic Nursing Practice, *7*(2), 34-38.

Maes, S., Verhoeven, C., Kittel F., & Scholten, H. (1998). Effects of a Dutch work-site wellness-health program: the Brabantia Project. American Journal of Public Health, *88*(7), 1037-1041.

Mokbel, M.F. (2007). Privacy In Location-Based Services: State-Of-The-Art And Research Directions. Proceedings of The International Conference On Mobile Data Management, Mannheim, Germany.

Maximilien, E.M., Grandison, T., Sun, T., Richardson, D., Guo, S., Liu, K. (2009). Enabling Privacy As a Fundamental Construct for Social Networks. Proceedings of the Workshop on Security and Privacy in Online Social Networking (SPOSN), Vancouver, Canada.

Nurmi, D., Wolski, R., Grzegorczyk, C., Obertelli, G., Soman, S., Youseff, L., & Zagorodnov, D. (2008). The Eucalyptus Open-Source Cloud-Computing System. Proceedings of Cloud Computing and Applications (CCA), Chicago, Illinois, USA. Retrieved April 4, 2011 from http://www.cca08.org/papers/Paper32-Daniel-Nurmi.pdf

Ozminkowski, R.J., Ling, D., Goetzel, R.Z., Bruno, J.A., Rutter, K.R., Isaac, F., & Wang, S. (2002). Long-Term Impact of Johnson & Johnson's Health & Wellness Program on Health Care Utilization and Expenditures. Journal of Occupational & Environmental Medicine, *44*(1), 21-29.

Ozturk, C., Zhang, Y., & Trappe, W. (2004). Source-Location Privacy In Energy-Constrained Sensor Network Routing. Proceedings of The 2nd ACM Workshop On Security Of Ad Hoc And Sensor Networks (pp. 88-93), New York, NY, USA.

Pearson, S., Shen, Y., & Mowbray, M. (2009). A Privacy Manager for Cloud Computing. Lecture Notes in Computer Science, *5931*, 90-106.

Pearson, S. (2009). Taking account of privacy when designing cloud computing services. Proceedings of the ICSE Workshop on Software Engineering Challenges of Cloud Computing (pp. 44-52).

Savolaine, J., & Granello, P.F. (2002). The Function of Meaning and Purpose for Individual Wellness. Journal of Humanistic Counseling, Education and Development, *41*, 178-189.

Sweeney, L. (2001). k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, *10*(5), 557-570.

Tibken S. (2010). Salesforce CEO touts mobility, social networking. Dow Jones Newswires. Retrieved April 4, 2011 from http://www.totaltele.com/view.aspx?C=4&ID=459551

Wang, J., Zhao, Y., Jiang, S., & Le, J. (2009). Providing privacy preserving in cloud computing. Proceedings of the International Conference on Test and Measurement (pp. 213-216).

Wilson, R.L., & Rosen, P.A. (2003). Protecting data through Perturbation Techniques: Impact on the knowledge discovery process. Journal of Database Management, *14*(2), 14-26.

Zane, R. (2009). IBM Research Collaborates with Leading Taiwanese Institutions to Deliver Wellness-Centric Healthcare Via Cloud Computing. Press Release. Retrieved April 4, 2011 from http://www-03.ibm.com/press/us/en/pressrelease/29086.wss

Zeng, L., Hsueh, P., Chang, H., Chung, C., & Huang R. (2010). GreenOlive: An Open Platform for Wellness Management Ecosystem. Proceedings of the IEEE/INFORMS International Conference on Service Operations and Logistics, and Informatics (SOLI), Beijing, China.

ADDITIONAL READING SECTION

Adam, N.R. & Wortmann, J.C. (1989). Security-Control Methods for Statistical Databases: A Comparative Study. ACM Computing Surveys, *21*.

Applebaum, P.S. (2000). Threats to the Confidentiality of Medical Records—No Place to Hide. JAMA, *283*(6), 795–796.

Bayardo R.J., & Agrawal, R. (2005). Data Privacy Through Optimal k-Anonymization. Proceedings of the 21st Int'l Conf. on Data Engineering (ICDE), Tokyo, Japan.

Chen, B., Lefevre, K., & Ramakrishnan, R. (2007). Privacy skyline: Privacy with multidimensional adversarial knowledge. Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB), University of Vienna, Austria.

Department of Health and Human Services. (1999). Standards for Privacy of Individually Identifiable Health Information. Federal Register, *64*(212).

Etzioni, A. (1999). Medical Records: Enhancing Privacy, Preserving the Common Good. Hastings Center Report (pp. 14–23), Mar–Apr 30.

Grandison, T., & Davis, J. (2007). The Impact of Industry Constraints on Model-Driven Data Disclosure Controls. Proceedings of the 1st International Workshop on Model-Based Trustworthy Health Information Systems, Nashville, Tennessee.

Lefevre, K., Agrawal, R., Ercegovac, V., Ramakrishnan, R., Xu, Y., DeWitt, D. (2004) Limiting Disclosure in Hippocratic Databases. Proceedings of the 30th Int'l Conf. on Very Large Databases, Toronto, Canada.

Loukides G. & Shao, J. (2008). Data Utility and Privacy Protection Trade-off in k-Anonymisation. Proceedings of the First International EDBT Workshop on Privacy and Anonymity in the Information Society (PAIS), Nantes, France.

Samarati, P. (2001). Protecting Respondents' Identities in Microdata Release. IEEE Transactions of Knowledge and Data Engineering, *13*(6), 1010-1027.

Savolaine, J., & Granello, P.F. (2002). The Function of Meaning and Purpose for Individual Wellness. Journal of Humanistic Counseling, Education and Development, *41*, 178-189.

Sweeney, L. (1997). Weaving Technology and Policy Together to Maintain Confidentiality. Journal of Law, Medicine & Ethics, *25*, 98–110.

KEY TERMS & DEFINITIONS

Application Programming Interfaces (API): A pre-determined set of functions that specify how programmers utilize the features of a software program (which can be a library, an application, an operating system, or a network device driver).

Ecosystem-as-a-Service (EaaS): An economic community formation model which produces goods and services of value to customers, who are themselves members of the ecosystem.

Data Perturbation: Techniques that are used to insert minor biases into databases, either directly on the data or on the output of query result.

Health Insurance Portability and Accountability Act (HIPAA): The HIPAA Privacy Rule provides federal protection for the confidentiality, integrity, and availability of Personal Health Information (PHI) held by covered entities. It gives patients an array of administrative, physical, and technical protections with respect to PHI and specifies rules for the disclosure of PHI needed for patient care or research purposes.

Health Information Technology for Economic and Clinical Health Act (HITECH): Integrated as part of the economic stimulus bill, American Recovery and Reinvestment Act of 2009, to encourage healthcare providers to use electronic record-keeping and ordering system.

Hypertext Transfer Protocol Secure (HTTPS): Is a combination of the Hypertext Transfer Protocol with the SSL/TLS (Secure Sockets Layer/ Transport Layer Security) protocol to provide encrypted communication and secure identification of a network web server.

Infrastructure-as-a-Service (IaaS): A provisioning model utilized by a platform operator to outsource its equipment to support third-party operations, including processes, storage, and networking components. It is typically operated on a pay-as-you-go basis.

Personal Health Information (PHI): PHI includes demographic information, medical history, test and laboratory results, insurance information and other data that is collected by a healthcare professional to identify an individual and determine appropriate care.

Personally Identifiable Information (PII): Data about an individual that could distinguish and trace an individual, such as name, age, email address, mailing address, telephone number, social security number, fingerprints, other biometric data, medical or financial information.

Publish-Subscribe Service (Pub-Sub): Pub-sub service provides communication channels among services to allow one service send a message on a particular topic, and all the other services that have subscribed to this topic to receive the message.

Secure Sockets Layer: a cryptographic protocol developed by Netscape for transmitting private documents via the Internet.

Software-as-a-Service (SaaS): An on-demand software distribution model that made available the applications hosted by a service provider to end users through web services in a service-oriented architecture. It is typically operated on a pay-as-you-go basis.

Wellness: the active process of becoming aware of and making choices toward a more successful existence.

Wellness Cloud (WC): A Wellness Cloud is an integrated, interconnected and intelligent platform for the ingestion, processing and management of wellness data that delivers services to multiple independent software vendors (ISVs), service providers, and other stakeholders.