MyHealthAvatar

A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information

Project acronym: MyHealthAvatar

Deliverable No. 6.1
Initial report on data collection utilities and evaluation

Grant agreement no: 600929

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	MyHealthAvatar
Project Full Name:	A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information
Deliverable No.:	D6.1
Document name:	Initial report on data collection utilities and evaluation
Nature (R, P, D, O)[1]	R
Dissemination Level (PU, PP, RE, CO)[2]	PU
Version:	1
Actual Submission Date:	20/02/2014
Editor: Institution: E-Mail:	Po Yang University of Bedfordshire po.yang@beds.ac.uk

ABSTRACT:
This deliverable presents the initial report on constructing a data collection utilities solution for creating MyHealthAvarter platform. The data collection utilities solution (DCU) in MHA investigated a number of data collection utilities which will collect information of individual patients without requiring a major effort from the patients themselves. This deliverable reviewed existing data collection utilities solution that holds the environmental information of the patients' life styles, diet, and other clinical related information by developing a number of tailored information extraction tools for the web and mobile apps. More specifically, it targeted information extraction from social network, which falls in the category of information extraction. This deliverable reported the detailed design and implementation of DCU solution in MHA platform. It looked into existing approaches and open sources, such as Fitbit, Moves and MyTracks, etc. Also, the health data extractions from social media have been implemented. An integrated data collection platform for users is built for retrieving third-party health information to MHA data repositories. It will support the users to include information such as: Personal information (contacts, providers, ID on social network, etc.), Sleep, Glucose/Blood sugar, Food, Hear rate and Mood. The future work would

[1] **R**=Report, **P**=Prototype, **D**=Demonstrator, **O**=Other

[2] **PU**=Public, **PP**=Restricted to other programme participants (including the Commission Services), **RE**=Restricted to a group specified by the consortium (including the Commission Services), **CO**=Confidential, only for members of the consortium

[2] **PU**=Public, **PP**=Restricted to other programme participants (including the Commission Services), **RE**=Restricted to a group specified by the consortium (including the Commission Services), **CO**=Confidential, only for members of the consortium (including the Commission Services)

focus on the extension of existing data collection platform to compatible more sensing data.

KEYWORD LIST:
Data collection utility, social network, healthcare platform

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n^{o} 600929.

MODIFICATION CONTROL			
Version	**Date**	**Status**	**Author**
0.1	13/01/2014	Draft	Po Yang Zhikun deng
0.2	01/02/2014	Pre-final version	Po Yang
1.0	25/02/2014	Final version	Po Yang

List of contributors

- Po Yang
- Zhikun Deng

Contents

1 EXECUTIVE SUMMARY 5
2 INTRODUCTION 6
2.1 PURPOSE OF THIS DOCUMENT 6
3 TECHNICAL INVESTIGATION 7
3.1 MOBILE APPS BASED HEALTH DATA COLLECTION 7
3.2 DEVICE BASED HEALTH DATA COLLECTION 10
3.3 HEALTH DATA MINING IN SOCIAL MEDIA 12
3.4 PLATFORM BASED HEALTH INFORMATION SHARING 13
3.5 DISCUSSION 14
4 DATA COLLECTION UTILITIES SOLUTION IN MHA 16
4.1 INTRODUCTION 16
4.2 DATA TYPE 16
4.3 ARCHITECTURE 17
4.4 EXPERIMENTAL TRIALS 18
5 IMPLEMENTATION 19
5.1 INTRODUCTION 19
5.2 DCU IMPLEMENTATION OVERVIEW -- SAAS 19
5.3 DCU IMPLEMENTATION OVERVIEW – OAUTH 1/2 20
5.4 DCU IMPLEMENTATION OVERVIEW – SPRING SOCIAL FRAMEWORK 23
5.5 MHA DCU IMPLEMENTATION OVERVIEW – THREE LAYERS STRUCTURE 24
5.6 FRONTEND WEB UI 26
5.7 DCU DATA POST PROCESSING ARCHITECTURE 28
5.8 CHALLENGES ..... **ERROR! BOOKMARK NOT DEFINED.**
6 CONCLUSION 29
6.1 SUMMARY OF CURRENT STAGE 29
6.2 CHALLENGES 29
6.3 FUTURE WORK 30
7 REFERENCES 31

1 Executive Summary

Personal health related data collection is of significance to success of lifetime European healthcare systems (MyHealthAvatar project). Differing in other traditional Electronic Medical Records (EMR) and Electronic Health Records (EHR) with the main purpose of clinic use, the collection of personal health related data are particularly aimed for citizens to self-manage their health information. Recently, the emergence of numerous wearable devices and mobile apps for health and fitness uses provides the opportunity to collect large-scale timely personal health information. However, due to the diversity and uncertainty of devices, the health data retrieved from these applications are usually inaccurate and stored in diverse formats both logically and physically. Healthcare data come from many resources and is delivered in many forms. It is desirable to deliver a uniformed data collection solution enabling citizens to record, store, operate, visualize and share their personal health information.

MyHealthAvatar (MHA) proposes a solution for access, collection and sharing of long term and consistent personal health status data through an integrated environment. The data collection utilities solution (DCU) in MHA investigated a number of data collection utilities which will collect information of individual patients without requiring a major effort from the patients themselves. In this deliverable we reviewed existing data collection utilities solution that holds the environmental information of the patients' life styles, diet, and other clinical related information by developing a number of tailored information extraction tools for the web and mobile apps. More specifically, we targeted information extraction from social network, which falls in the category of information extraction. This deliverable reported the detailed design and implementation of DCU solution in MHA platform. We looked into existing approaches and open sources, such as Fitbit, Moves and MyTracks, etc. Also, the health data extractions from social media have been implemented. An integrated data collection platform for users is built for retrieving third-party health information to MHA data repositories. It will support the users to include information such as: Personal information (contacts, providers, ID on social network, etc.), Sleep, Glucose/Blood sugar, Food, Hear rate and Mood.

This document describes all the activities planned for Task6.1 in respect to the data collection utilities and linking to social media. If necessary this document will be dynamically updated in the course of the project to reflect project's findings, and will extend the list of technologies, standards, protocols, etc.

2 Introduction

This document includes all related activities that form the core architecture of the data collection utilities solution to support MyHealthAvatar (MHA). We will discuss the latest health data collection technology in both industrial and research fields, including mobile apps based health data collection method, device based health data collection method and health data mining in social media. Though deep investigations of these techniques, a feasible data collection utilities solution for MHA platform would be proposed and implemented. This solution would be built on the existing market-available techniques and support the health data retrieving from multiple devices, mobile apps on different platforms, including iOS and Android. Relying on this data collection solution, MyHealthAvatar platform can be maximally compatible with the existing health and fitness mobile apps. However, some issues with respect to the data storage and security have to be considered in this document.

2.1 Purpose of this document

Currently, there are a variety of health and fitness data collection mobile apps and devices in the market, such as Fitbit, Endomondo and Nike *Fuelband*. While most of them are particularly targeted at the fitness group, the widely spread of these wearable products is generating a new revolution of traditional healthcare system. The goal of MHA project is to deliver a platform for general public to access integrative models and analysis tools, utilizing resources already created by the VPH community. With MHA platform, people can easily gather, store, use and share their personal health information for themselves and their family. In order to reach the goal, a feasibly effective data collection utilities solution is of importance to success of the MHA platform.

The aim of this deliverable is to review and analyse the available health data collection techniques and justify the most feasible solution for MHA platform. In order to reach this goal, we would firstly investigate the possible extracted type of health data and the relevant existing techniques to retrieve them. Then, considering the availability of API and the support of platform, an integrated data collection utilities solution is designed and implemented for MHA. Finally, the discussion of challenging issues and future work are addressed in this deliverable.

3 Technical Investigation

In the sections below we reviewed the existing health data collection techniques, including mobile apps based health data collection, device based health data collection, health data mining in social medial and platform based health information sharing. The data collection utilities solution in MyHealthAvatar project would attempt to utilize the above existing techniques and improve them further as a feasible data collection utilities solution.

3.1 Mobile apps based health data collection

Mobile applications recently are turning out to be a great source of user empowerment. The smart phone applications have engulfed everywhere in personal and professional life. This phenomenon has significantly impacted healthcare sector. In this section, we reviewed some popular mobile applications for health data collection. The most well-known mobile apps are based on GPS information for tracking user moving activities. The type of health data includes the location, distance, speed and other manual recorded health data.

3.1.1 **Endomondo**

Endomondo is a popular GPS based mobile application for tracking route, distance, duration, split times and calorie consumption, as shown in Figure 3.1. It offers a full history with previous workouts and statistics, as well as a localized route map for each work out. Also, this app supports integration with different other device based health information input, like heart rate monitors.

Figure.3.1 The user interface of Endomondo

One key feature of Endomondo is that it can incorporate community and allow users to challenge friends and share results. By integrating the feature of social networking, users can see the activities and performance of their friends and colleagues on Endomondo. This feature could maximally attract new users to join the community to share their journey and commiserate. Currently, Endomondo supports several mobile platforms, including iOS, Android, RIM, Windows Phone, Symbian, Window Mobile and Java. The app can be downloaded for free and a premium version is also available for $4.99. The main drawback of Endomondo is short battery longevity since GPS sensor is quite power-consumption.

3.1.2 Google MyTracks

Similar to Endomondo, Google MyTracks is also based on the use of GPS to record users' path, speed, distance and elevation while they are walk, run, and bike or do any activities outside as shown in Figure.3.2.

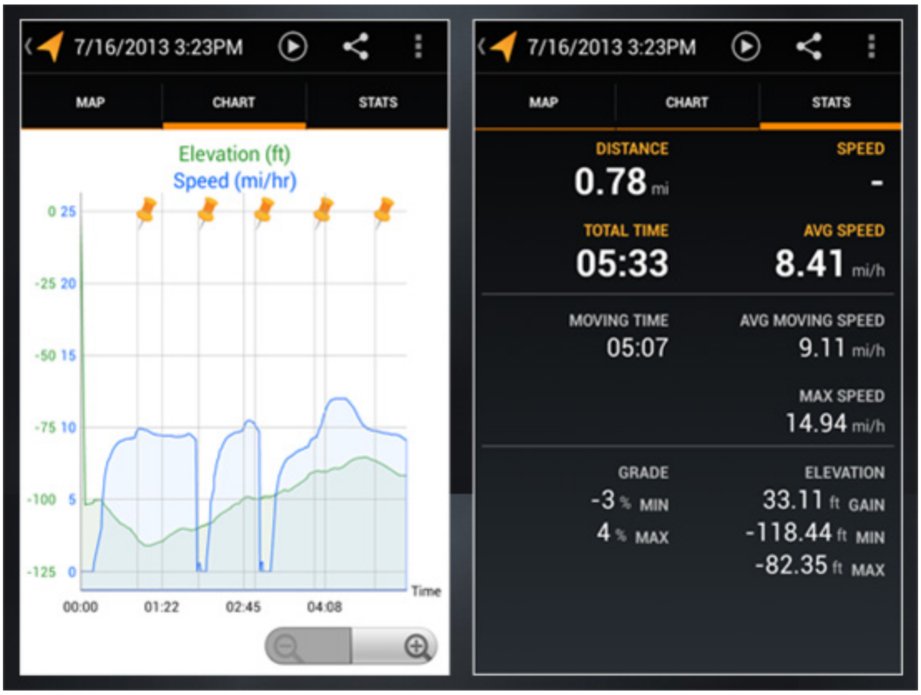

Figure.3.2 The interface of My Tracks

After user recording their location information, they can view their data live, annotate their path and hear periodic voice announcements of their progress. Compared to Endomondo, My Tracks is fully supported by Google with a comprehensive API documentation. While it is only available on Android platform, it is quite popularly used and further developed by users. This app can also be downloaded for free. However, it has a similar drawback of battery life on turning on GPS receiver.

3.1.3 Cardiio

Cardiio is a simple touch-free heart rate monitoring mobile application, as shown in Figure.3.3. It can measure heart rate from a distance by simply having users' face straight into the front camera of smart phone for few seconds. Compared to other heart rate monitoring approaches, Cardiio is a completely device-free solution. The fundamental technology is that every time human heart beats, more blood is pumped into their face. This slightly increase in blood volume causes more light to be absorbed, and hence less light is reflected from their face. The mobile application is developed by a cutting-edge research and science team at the MIT Media Lab.

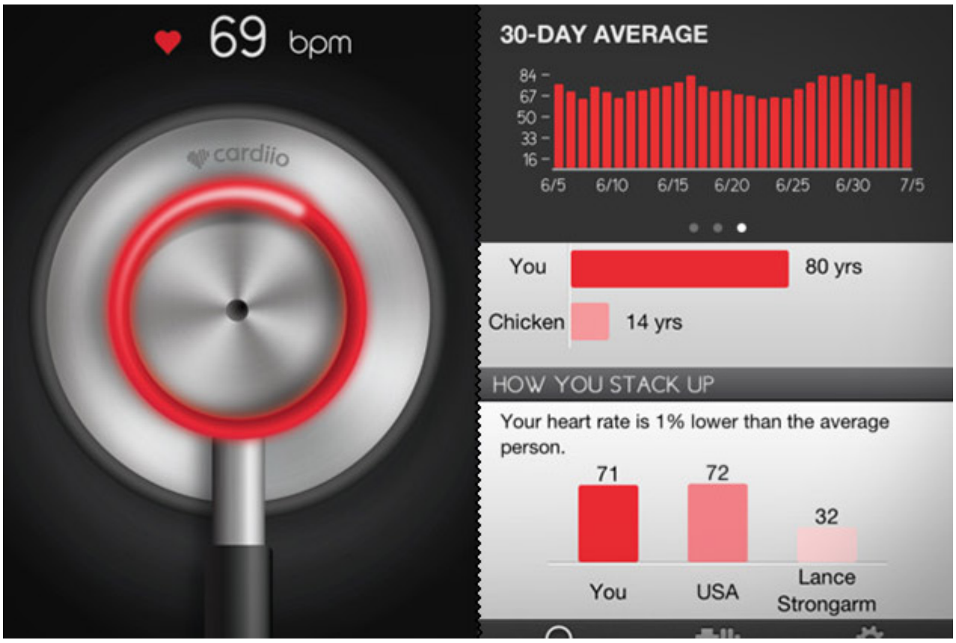

Figure.3.3 The interface of Cardiio

The limitations of Cardiio is that the accuracy is not high and only available on iOS platform. There are no accessible API support to carry out the further development.

3.1.4 EmotionSense

Emotion sense is an Android application that lets user explore how they mood relates to the data that their smart phone can invisibly capture. In this mobile application, the experience sampling method is used to collect longitudinal survey data from participants. The smart phone sensors are used to augment the context-awareness of sampling strategies. As shown in Table.3.4, the mobile application collects data by asking users to complete short survey about their mood at difference times of the day. When a survey is read to be answered, an icon will appear in the notification bar. The user can see their aggregated responses on the Emotion Grid screen by clicking on this notification. The more answers are replied by users; the more detailed users' mood would be demonstrated.

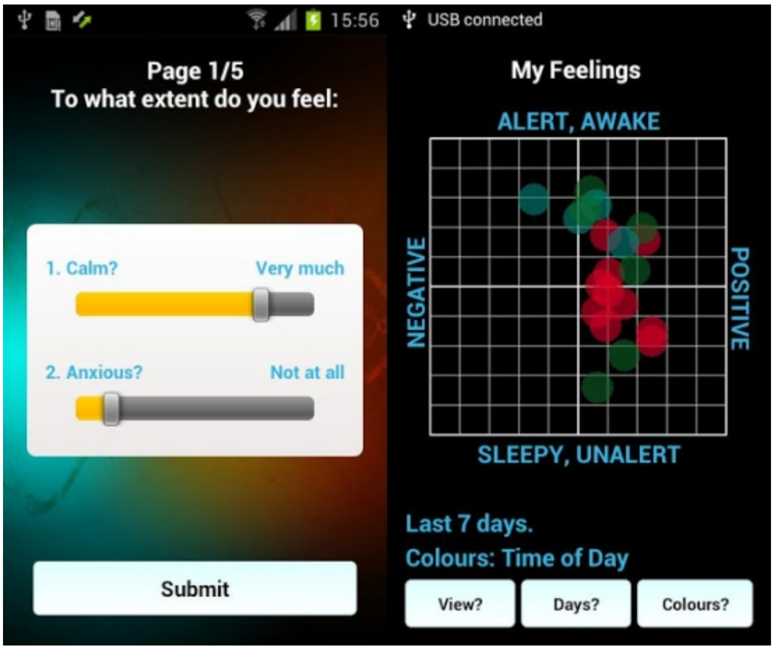

Figure.3.4 The interface of EmotionSense

The limitations of EmotionSense is that it is a purely research-oriented application. Current available version can be only downloaded on iOS platform. There is no accessible API support to carry out the further development. Meanwhile, the emotion data is collected though manual input but not automatic retrieving by users.

Apart from the above typical mobile applications, a variety of healthcare applications are available to be used in the market. Most of them are free-download and with similar functionalities.

3.2 Device based health data collection

Device based health data collection used to refer to using some medical devices particularly to collect some medical data, like blood pressure, heart ratio, glucose, etc. Here, we mainly consider a new direction of using latest wearable computing technologies for collecting life-logging health data, such as wearable twists.

3.2.1 **Wearable device based physical activity data collection**

Personal physical activity data has already been shown to be effective in various health and fitness applications. In particular, prior work has shown that wearable sensors can benefit individual patient health [1] and individual personal fitness [2]. Early work examined data collected from specialized or research-grade accelerometer-based devices [3]. Seeing the benefits of this technology, several companies have now produced wearable activity monitoring devices at price levels that are attractive to everyday consumers. We listed them as below:

Fitbit [4] has four wearable devices on the market, *Ultra*, *One*, *Zip*, and the *Flex* (recently released). With the exception of the *Flex*, all are made to be discreet and wearable either on trousers or shirts. The *Flex* is worn on the wrist. The devices record steps taken, distance travelled, and calories expended. The *Ultra* and *One* have an altimeter that allows the counting of the number of floors walked up. These devices communicate with a host computer using Bluetooth that sends their data directly to a user's account on the *Fitbit* website.

Nike+ Fuelband [5] (the plus is pronounced) is worn on the wrist and records calories, steps, distance, and Nike's own unit of activity terms "Nike Fuel". Accumulative amounts of each item can be retrieved through a small display on the device. The device connects via USB to a host machine which syncs the data to a user's account on the Nike+ website.

Nike+ Sportsband/Motion This device is worn on the wrist and works with the *Nike+ Motion* sensor which is worn on an individual's shoe [6]. This device only records distance, which can be uploaded via USB to Nike website. The Nike+ *Motion* sensor can also be used with a Nike watch and iPhone App.

Jawbone Up It is worth mentioning the *Jawbone Up* [7], which is the other high-profile consumer level activity device launched in 2011. The device, however, was pulled off the market after its initial launch due to faulty batteries and leaks. The product has only been recently re-launched in Nov 2012, while we were performing the study. As a result, we have omitted this from our study. The *Jawbone Up* provides steps, distance, calories. Currently the *Jawbone up* can only be used with mobile device, drivers for laptop and PCs are not provided.

GPS Watches We have omitted GPS-based watches, such as those made by *Garmin* and *Timex*, because they still represent high-end devices with costs typically exceeding USD$200. Moreover, these devices are not intended to record daily activity, but are used while engaging in specific activities, e.g. running.

Regarding the accuracy of above five wearable devices, it is reported [8] that the *Fitbit* is capable of delivering a very low error of around 1% for step recording. While distance error was higher, it was highly correlated with steps. Other devices showed significantly more error, with *Nike+ Fuelband* at almost 8% error for steps and *Sportsband* with over 10% for distance - both with significant variations. *Nike+ Fuelband* only provided distance as a daily accumulation on their website and could not be accurately measured per 400m lap walked.

As the scenario reflects, Fitbit alleviates the burden of self-reporting by automatically tracking a person's daily step count. However, it cannot intervene to remind users that they have been sitting for too long and should take a short walking break. Another positive aspect of the *Fitbit* device is its API. *Fitbit* has an API that allows its information to be extracted with a per-minute step readings (the information is obtained by connecting to the webpage and not the Fitbit itself). While a 3rd party API is available to get data from the *Nike+ Sportsband* webpage, the API for *Fuelband* is currently only open to select developers. Another nice benefit was that *Fitbit* was the Bluetooth connection that made it easier than the *Nike Fuelband* to collect data. We do note, that one downside of the *Fitbit* was that it was relatively easy to lose the device, compared to the Nike Fuelband. A Fitbit's was lost by one participant who did not notice that it had fallen off over the course of the day.

3.2.2 Medical device based healthcare data collection

Apart from wearable sensor based personal physical activity collection, many medical devices are also able to be automatically connected to smart phone. For instance, as shown in Figure.3.5, the latest iHealth blood pressure check and LifeScan devices can simultaneously transfer the measure of health data (blood pressure and glucose) to smart phone; further upload them to the personal health management platform.

Figure.3.5 iHealth blood pressure check and LifeScan

The health data collection by using medical device is mostly manual-mode, which is probably a weakness in comparison to the automatic mode of most wearable sensor based health data collection. However, it has been improved significantly, particularly considering that some data has to be hardly collected with non-obstructive way, like glucose.

3.3 Health data mining in social media

The popularity of social media allows the websites like Twitter, Facebook or PatientsLikeMe become communication hubs where people tell personal stories and share life experience. This hub contains a large volume of potential personal health information. The use of date mining technique in the exploration of personal health information from these social networking websites has become a hot research area. The key problem is how to search and extract the keyword from a large volume of plain text. In this section, we would review some papers on how to retrieve the health information from Twitter and Facebook.

3.3.1 Twitter

In paper [9], it explore the potential of mining twitter data to provide a tool for public health specialists and government decision makers to gauge the degree of concern (DOC) expressed in the tweets of Twitter users under the impact of diseases. The original motivation is that the early discovery of public health concerns can help governments to make timely decisions to refute rumours, and thus prevent potential social crises such as Chinese panic-buying of salt.

In order to measure the DOC in tweet datasets, paper [9] developed a classification approach to analyse sentiments in disease-related tweets and automatically classify personal tweets into negative tweets or non-negative (neutral) tweets. This approach involves a two-step method. It first distinguishes personal from news (non-personal) tweets. Since Twitter users broadcast news tweets released by online media organizations, possibly redistributed by Twitter users using 're-tweets,' it considers the news tweets as non-personal as opposed to personal tweets posted by individual Twitter users. It refers to the former as news tweets and the latter as personal tweets. In the second stage, the sentiment analysis is applied to personal tweets to identify personal negative tweets, as opposed to neutral (non-negative) tweets. Although news tweets may also contain concerns about a certain disease, they tend not to reflect the direct emotional impact of that disease on people. A person re-tweeting (like forwarding an email) a news message about a disease is most likely not directly affected by it, while a user sending out a personal tweet with emotional expressions might be directly affected.

Also, in paper [10], two semantic-based methods for mining personal health information from Twitter are introduced. One method uses WordNet as a source of health-related knowledge, another terms of personal relationships.

3.3.2 **Facebook**

In paper [11], it reports a study of analysing emotion on the wall messages of Facebook. In their studies, three databases are built to store related emotion terms and symbols for three resources of information on Wall messages of Facebook in Chinese, English and Emoticon. The wall messages of the friends of a user can be analysed by the developed Facebook application. The application can rank the friends of the user based on the attention index developed in this study.

Based on their study, paper [11] developed a system called Friends_Smile on Facebook, which can judge the emotions of selected friends and calculate their degree of attention needed from their post messages in the past one week. The experimental results demonstrate that the system is useful to be a pioneer study in emotion extraction though social networking.

Regarding the above review, it appears that the social networking has a huge potential to enhance the quality of personal health information collected from other utilities. However, due to the legal issue of data protection, these studies are still in the pure research period, no commercial applications have been successfully implemented.

3.4 Platform based health information sharing

The platform based health information sharing can be traced to electronic healthcare systems for decades with traditional healthcare providers such as hospitals, medical centers, laboratories and doctors' offices. Due to the great evolution of internet technology, the web-enabled healthcare service is emerging as a new healthcare delivery trend. In these web based healthcare platforms, they provide a multi-functional server for users to store, manage and visualize health data from various third party devices.

3.4.1 Microsoft HealthVault

The idea of Microsoft HealthVault is to enable user to gather, store, use and share personal health information though many medical devices. Currently, HealthVault enables a connected ecosystem including more than 300 applications and more than 80 connected health and fitness devices. It is a cloud-based platform to put people in control of their health data. An important feature of HealthVault is that it offers a privacy and security enhanced foundation on which a broad ecosystem of solution providers, device manufacturers and developers build innovative new health and wellness management solutions.

3.4.2 MyFitnessCompanion

MyFitnessCompanion is another healthcare platform for users to manage their personal health data, including metrics like weight, heart rate & HRV, Blood Pressure, Food intake, blood glucose, insulin, asthma, etc. The functionalities are highly similar to Microsoft HealthVault, even able to synchronize with Microsoft HealthVault and other health services. The main feature of MyFitnessCompanion is that it has a real-time visualization mode, which can keep track and visualize all users' measurement with graph on time and share these graphs with others.

Apart from the above two platforms, there are also other similar available healthcare platforms. The general trend is that most mobile apps and devices providers have built up their own server to store health information. Therefore, the device connectivity and information sharing between them is a critical issue.

3.5 Discussion

Regard the above technique investigation, it appears that personal user activities based health data is capable of being popularly collected by mobile apps. In designing a data collection uncivilities solution for MHA platform, the primary issue of consideration is the possible identification of data type. From the section 3.1-3.2, most data related to physical activities such as position, steps, distance and location can be automatic detected and tracked due to the mature GPS and sensor technology. These physical data can be the first identified data type in MHA platform.

Some data types like Glucose and Blood sugar are hardly collected regarding the need of extra equipment. These data are probably particularly useful for some patients, but not quite necessary to general publics. They would be not included in our first version of data collection utilities solution, but can be included in the further version.

Some data types like food; sleep and mood can be collected though mobile apps, but not be popular used by public. The reason is that they normally need a manual record. Also, heart rate is possibly collected without accessing extra equipment but the accuracy is not reliable. These data are supported into the first version of data collection utilities solution, but the quality and accuracy of these data have to be improved when they are used for prediction or diagnosis in MHA platform.

Another important issue is the availability of API and support of both iOS and Android system. For most of mobile apps or wearable device, they actually support both iOS and Android system. Google, Fitbit and Moves are all able to provide APIs for third party to retrieve health data from their server. But the preliminary requirement is that users have to register into their server firstly.

For the health data from social media, there are some concerns about user data protection and legal consideration. So far there are no commercial applications or platforms practically implementing the extraction. Also, the deep analysis of plain text from social media requires some ontology and semantic techniques. Consequently, in this deliverable, we would design and implement the linking to Twitter and Facebook, but not practically release it until the further clear clarification of data protection and legal issues.

4 Data Collection Utilities Solution in MHA

4.1 *Introduction*

In the section below we proposed an architecture design of data collection utilities solution in MHA platform. Firstly, it identifies the data type supported by MHA. Then it would describe the initial idea of architecture of data collection utilities solution in MHA platform. Finally, this section reports some experimental trials with Fitbit and Google MyTracks.

4.2 *Data type*

In the section, we listed the identifiable data type of DCU solution in MHA platform in Table 4.1. It includes the activities, sleep, Glucose/Blood sugar, food, heart rate and Mood.

Data Type	Apps	Add Device	Support platform	Availability of API
Activities	My Tracks	×	Android	√
	Fitbit Flex	√	iOS/Android	√
Sleep	Fitbit Flex	√	iOS/Android	√
Glucose/Blood sugar	LifeScan	√	Not sure	×
	Glucose	√	Not sure	×
Food	Fooducate	×	iOS/Android	√
	Fitbit Flex	√	iOS/Android	√
Heart rate	Cardiio	×	Android	×
	Cardiograph	×	Android	×
Mood	EmotionSense	×	Android	×

Table. 4.1 Data type of DCU solution in MHA platform

4.3 Architecture

In the sections below we proposed an architecture design of data collection utilities solution in MHA project. The following graph Figure.4.1 demonstrates the overview of DCU in MHA architecture. DCU is designed in a way that it works with various health related data sources including but not limited to: wearable devices, mobile application, legacy devices and classic datasets. MHA not only collects useful information of users in a lifetime style, it also provides this information to users' other application / services through its API.

MHA acts as a central hub which interacts with other data services automatically or semi-automatically with users' manual assistant. Different data sources can be categories to:

- **Classic Data Source**: this is the data which users can feed in MHA DCU by uploading files, which includes hospital records, clinical trial spreadsheet, CT scan images.
- **Legacy Device**: the data from legacy devices are displayed directly to end-users without the ability to export electronically format. DCU provide easy to use form for user to manually input the data with timestamp to MHA.
- **Mobile Application**: DCU has adapter and interface which interact with popular health related mobile applications, this allows the data automatically collected and stored to MHA central database.
- **Wearable Device**: modern health monitoring device like Fitbit and Nike FuelBand+ SE. if the device provide API for DCU to interact with, the users' data will be collected from the devices' server through their defined security mechanism.
- **Social Network**: User's status update to social networks could include valuable information related to user's location, activity, food, etc. The health related information can be retrieved by DCU for analysis and pass to MHA.

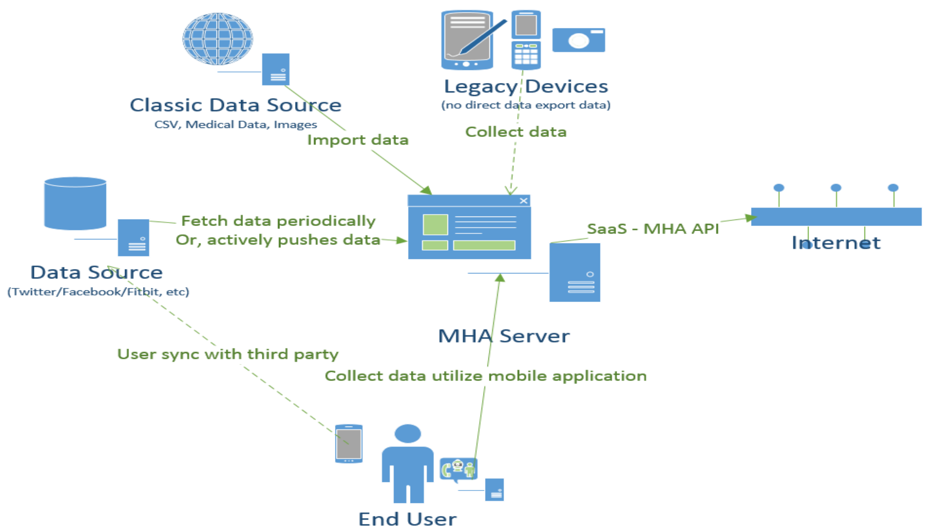

Figure.4.1 Data collection utilities solution in MHA

4.4 Experimental trials

In the sections below we proposed an architecture design of data collection utilities solution in MHA project. Fitbit devices, MyTracks and Moves application is evaluated for about five month since the early stage of the development process. Also we have registered account with popular social networks including Twitter, Facebook and Google+, for the purpose of interact with them to retrieve user's information.

DCU module works with all above mentioned services, and able to collect testers data without any problem. With the unified web interface, tester is able to go through similar configuration process for linking accounts and devices to MHA. Testers gives permission on data and authorize DCU to retrieve the data on behalf of them, through this mechanism, DCU is able to retrieve tester's profiles of different server, tester's friend list (if available), tester's manual status update, and tester's synchronized data from devices.

Collected data is then categories and post-processed before feed into MHA main data store, this is to only preserve information MHA interested in and clean the data for further analyses and visualization. All testers generally find it is very user friendly to use as well as work as expected, the combined data gives them more benefit over login to different servers and see data as scattered format.

We have limited resources and data collected so far is relatively limited, lots of machine learning related algorithm is pending on further collection of data from larger group of testers and users.

5 Implementation

5.1 Introduction

The Data Collection Utility (DCU) project of MyHealthAvatar (MHA) is a platform utility for user to gather, store health information in a style of life logging. DCU's data collection enables user to establish connections with Software-as-a-Service (SaaS) providers, to retrieve data on behalf of the user automatically. For other type of data formats, DCU has created generic data collection interfaces which allow DCU to collect data from various sources, which includes user's own input, devices without public APIs and some datasets exported from other applications (adapter might be required depends on data itself).

DCU module is itself a web based application which support user to interact from various devices and operating system, and all modern browsers of major platforms are supported. In order for DCU to automatically retrieve data on behalf of user, the user will be required to configure the devices and services through DCU's web UI. The UI is designed to be flexible and friendly for both desktop and mobile browsers, and the manual input option is designed in a way which takes into account intuitive for daily usage.

Following sections will mainly detail the implementation of DCU from a technique prospect, technical terminology and abbreviations are briefly explained on their first appearance, and short jargons will then be used where appropriate.

5.2 DCU Implementation Overview -- SaaS

The DCU mainly targets the services which provider mechanism to retrieve users' data with users' explicit authorization. These includes social networking services (e.g. Facebook, Twitter), health monitoring applications (e.g. MyTracks, Moves), and devices' public web API (e.g. Fitbit, Nike FuelBand+ SE). Over half a billion of world's internet users have flocked to social networking services to keep frequent update of their location, living and health conditions, DCU make uses of this data in a way which enhances MHA's overall functionality.

All targeted services have similar characteristics under their surface, these services can be simply treated as software applications that gather, store, process and share users' informations. Just as many existing applications, we abstract the services as users' data store which have users' credential management (sign up, sign in), users' activities record and other health information. The main advantages of chosen services are that the data they collect represent some facet of user' life log and health information, at the same time, these services are more than willing to share users' data with other services includes MHA (as long as the user gives permission to do so on his/her own data). This makes these SaaS great targets for MHA DCU to collect health data from.

The data collection from SaaS is a three-way conversation between:

1. A service provider (e.g. Fitbit server)

2. A service consumer (in our case: MHA DCU)

3. A user who holds accounts on both provider and consumer

Some of the service providers have sophisticated functionalities which stores lots of different data of their users, many of the data is not relevant to MHA. All interactions between DCU and service providers are scoped to the context of the user's health related data on the service provider side, and MHA will only retrieve the data which user specifically gives permission to.

Explain by Example:

To explain the process by example, assume Paul has account with Twitter and Fitbit (service providers). Paul wears Fitbit hand band everyday which records his activities, food and sleep data, at the same time Paul uses Twitter to update his feeling, travelling locations and sports related data. When Paul would like to see his daily activities, he needs to log on Fitbit server and see how many steps he has walked, how many calories he has burnt on a certain day. If he would also want to find out where he has been on that day, he would need to login his Twitter account and see if there is any location related tweet there.

On their own, Twitter and Fitbit provide great facilities for Paul to log his own feeling and travelling locations as well as his daily activities. However as someone who advocate to healthy living and well-being, it would be much better if Paul can sign in one place and view all health related information, even greater that the data are combined and presented to him in a much more meaningful way. When he tweet a message about his travelling or feeling, this data is automatically picked up, tagged and stored by DCU. When his Fitbit synchronize his steps with the Fitbit server, this information is also collected to form part of his well-being data by DCU. DCU makes MHA a perfect solution for Paul to use, and make his health monitoring process a lot more intuitive and efficient.

MHA DCU does not only data collection and tagging, it also provides additional mechanism to assist user to input health and living related information in a very user-friendly way. From users' perspective, MHA DCU provides very desired functionalities which bring together otherwise scattered health information in one location. MHA will also provide cutting edge visualization of the information to user which make sense and helpful.

The implementation of DCU involves the communication with service providers through standard security protocols. OAuth 1 and 2 are the most adopted and applied security standard for MHA to interact with service provider on behalf of the user. MHA take user information and security seriously, and following section will talk about security protocols used by MHA DCU.

5.3 DCU Implementation Overview – OAuth 1/2

OAuth is an open standard authorization protocol, which provides a mechanism for service consumer to access service provider's resources on behalf of the resource owner (e.g. the end-user). OAuth provide a process for end-users to authorize service consumer access to their data on service provider's server without sharing their credentials (username/password), which is normally implemented using browser redirections.

OAuth 1 and 1.0a is the more complex version of OAuth protocols, and it is used by Twitter, LinkedIn, and Fitbit etc. OAuth 1.0a is considered as more secure by some people; however it also requires significantly more computational power on both service provider and consumer. The workflow of OAuth 1 is illustrated as follows:

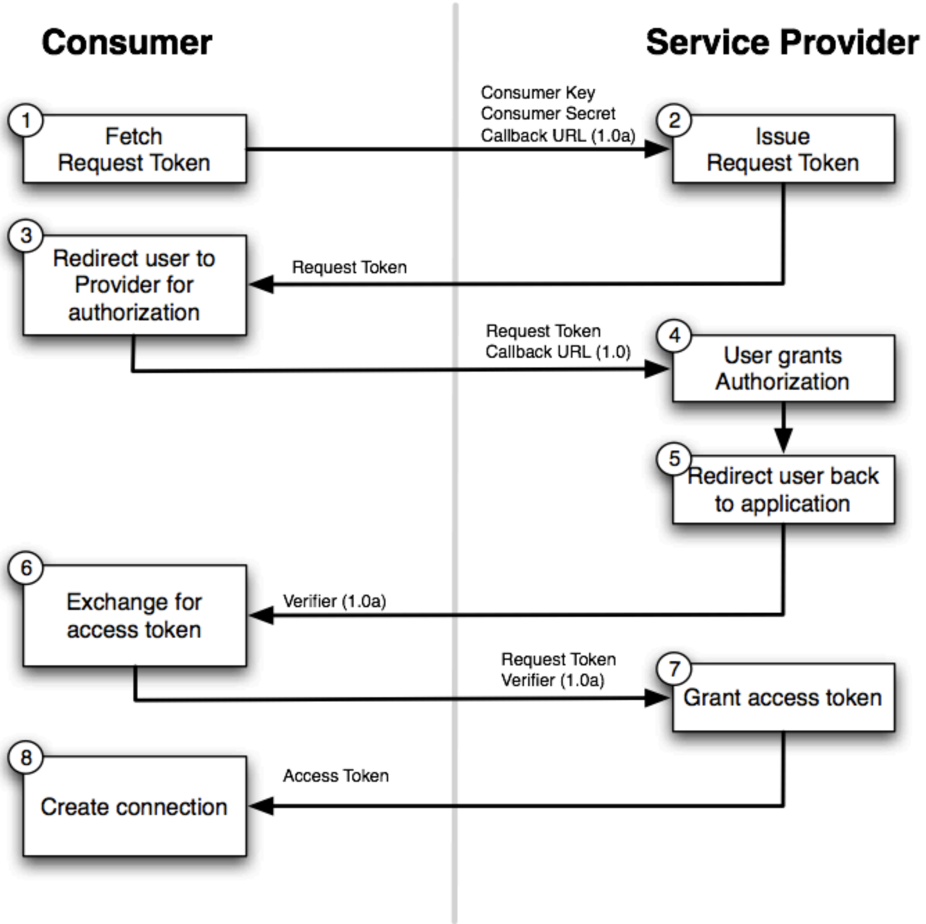

Figure 5.1 OAuth 1.0 Authorization Flow

1. The flow starts with the service consumer asking for a request token from service provider. The purpose of the request token is to obtain user approval and it can only be used to obtain an access token. In OAuth 1.0a, the consumer callback URL is passed to the provider when asking for a request token.

2. The service provider issues a request token to the consumer.

3. The service consumer application redirects the user to the service provider's authorization page, passing the request token as a parameter. In OAuth 1.0, the callback URL is also passed as a parameter in this step.

4. The service provider prompts the user to authorize the consumer application and the user agrees to the request.

5. The service provider redirects the user's browser back to the application (via the callback URL). In OAuth 1.0a, this redirect includes a verifier code as a parameter. At this point, the request token is authorized.

6. The application exchanges the authorized request token (including the verifier in OAuth 1.0a) for an access token.

7. The service provider issues an access token to the consumer. The "dance" is now complete.

8. The application uses the access token to establish a connection between the local user account and the external provider account. With the connection established, the application can now obtain a reference to the Service API and invoke the provider on behalf of the user.

OAuth 2 is becoming a preferred authorization protocol, and is now used by major providers such as Facebook and Github. The workflow of OAuth 2 looks like below:

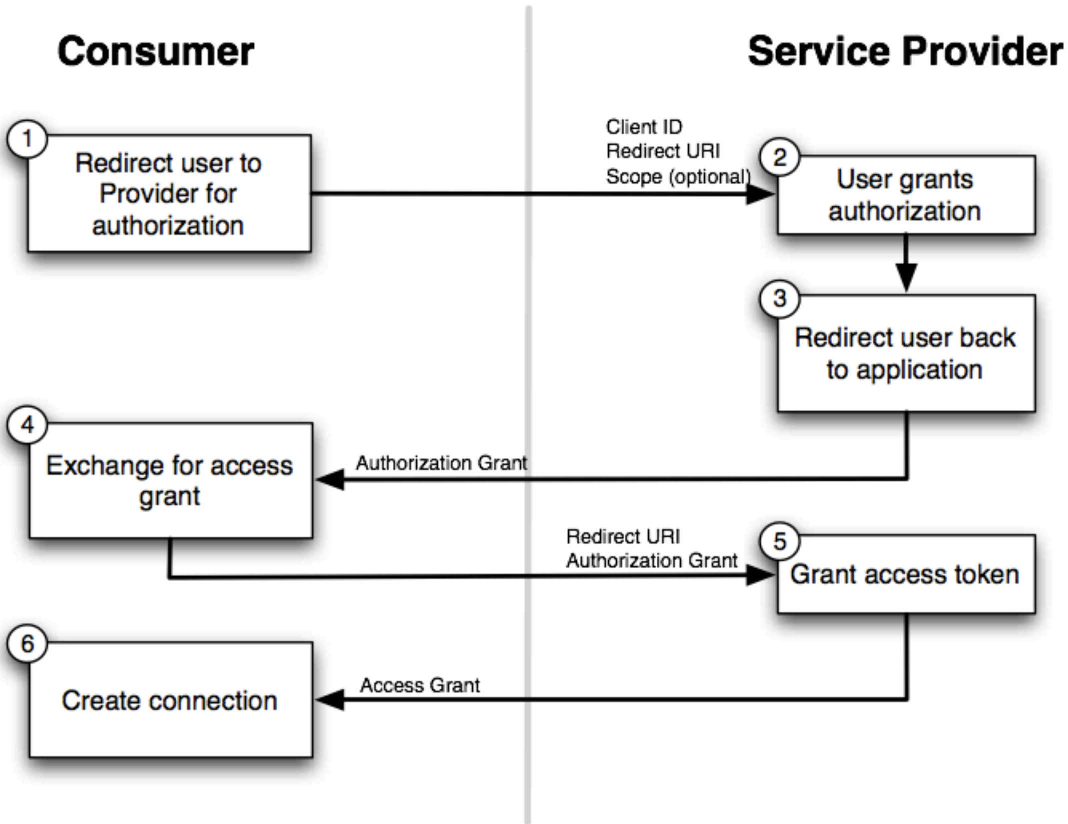

Figure 5.2 OAuth 2.0 Authorization Flow

As you can see, there is a back-and-forth conversation that takes place between the application and the service provider to grant the application access to the provider account.

This exchange, commonly known as the "OAuth Dance", follows these steps:

1. The flow starts by the service consumer application redirecting the user to the service provider's authorization URL. Here the provider displays a web page asking the user if he or she wishes to grant the application access to read and update their data.

2. The user agrees to grant the service consumer application access.

3. The service provider redirects the user back to the application (via the redirect URI), passing an authorization code as a parameter.

4. The application exchanges the authorization code for an access grant.

5. The service provider issues the access grant to the application. The grant includes an access token and a refresh token. One receipt of these tokens, the "OAuth dance" is complete.

6. The application uses the AccessGrant to establish a connection between the local user account and the external provider account. With the connection established, the application can now obtain a reference to the Service API and invoke the provider on behalf of the user.

MHA also implement in a way that it is a OAuth 2 service provider, which means that other application/service can interact with MHA's API on behalf of user (with user's authorization).

5.4 DCU Implementation Overview – Spring Social Framework

Java is the main programming language used by MHA main project, which is also the server language choice of MHA DCU module. Java runs on all major platforms include Windows, Linux and Mac OS, which makes Java very developer friendly from various OS backgrounds. Module based on Java is generally considered as more portable between OS, and does not bind the implementation to a specific platform. The portability benefits the development, test and publish process as developers generally use Windows and Mac OS while MHA serer are Linux boxes.

There are several popular Java OAuth implementations to choose from:

- Spring Social (http://projects.spring.io/spring-security/)
- OAuth Signpost (https://code.google.com/p/oauth-signpost/)
- Scribe (https://github.com/fernandezpablo85/scribe-java)

They are all open-source and free to use, which have their own advantages and disadvantages. The DCU's choice of library is based on the existing technology stack used by MHA, which is the Spring Framework. Choose Spring Social allows the team to implement in way which sticks to Spring Framework's convention and smoothly integrate with existing code base to avoid any potential conflicts. It can be overviewed as following graph:

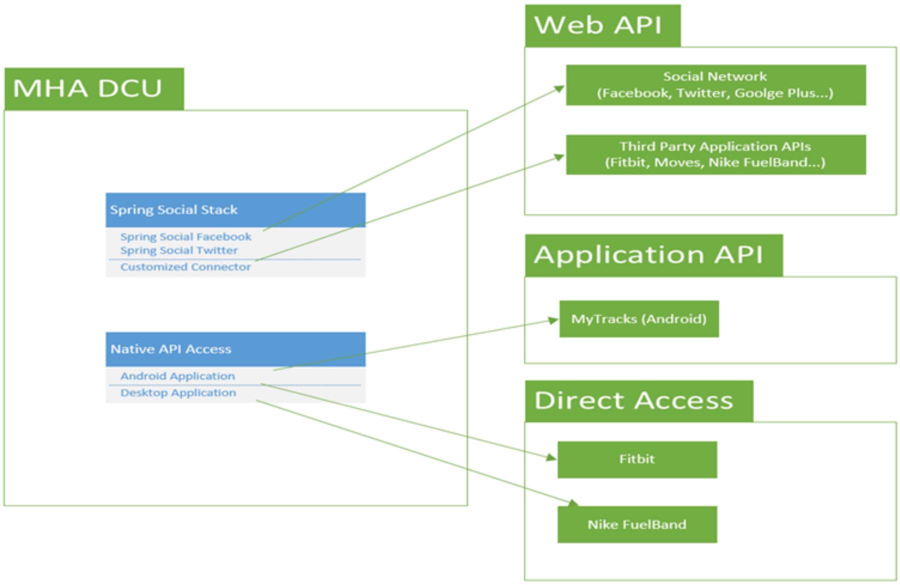

Figure 5.3 Overview of DCU Spring framework

DCU take advantage of Spring Framework's powerful libraries and favours Spring Framework's flavour of implementations when possible. Spring Security gives DCU a solid overall security framework which manages users credential and assist user's log in/out process. With Spring JdbcTemplate, DCU is able to efficiently store and retrieve the user's security tokens from other service providers and the details of database are abstracted nicely. Spring RestTemplate allows DCU to wrap service providers REST service in a consistent manner. Spring Web MVC provides a nice web framework to generate the frontend UI and handle user's requests.

5.5 MHA DCU Implementation Overview – Three Layers Structure

DCU module is not only a mechanism to interact with service providers and retrieve users' data; it also has its own friendly interface for user to control the authorization and permissions. More and more people nowadays uses mobile phone to access website and web applications, DCU take care of mobile users in its design, and responsive design pattern is applied to give DCU this flexibility to support standard and mobile browsers.

You can see how the same HTML page UI styled differently for different screen resolution, which not only allows information be displayed clearly to user, also gives user a much more touchscreen friendly experience.

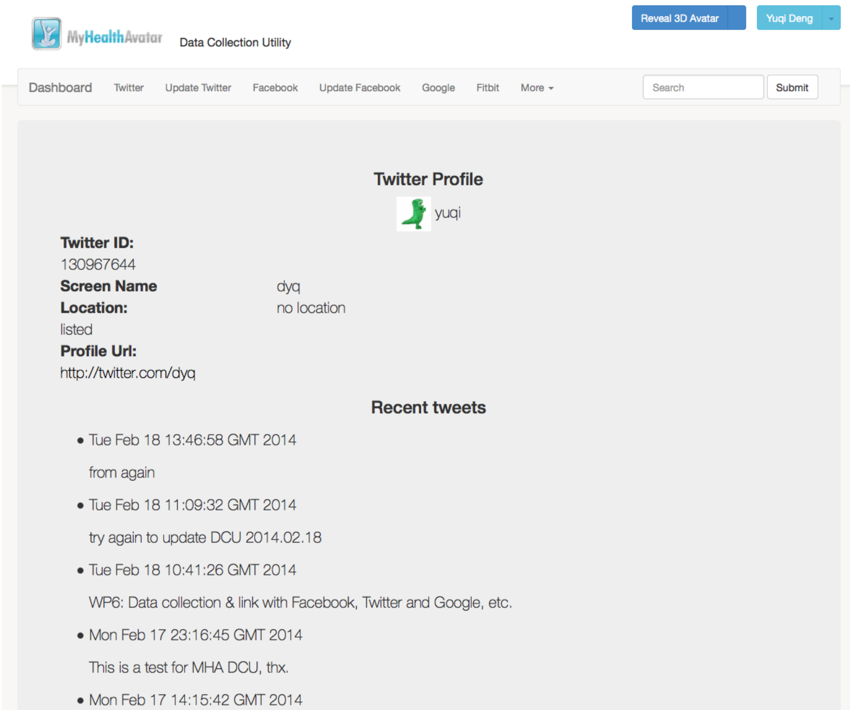

Figure 5.4 DCU UI from a standard browser

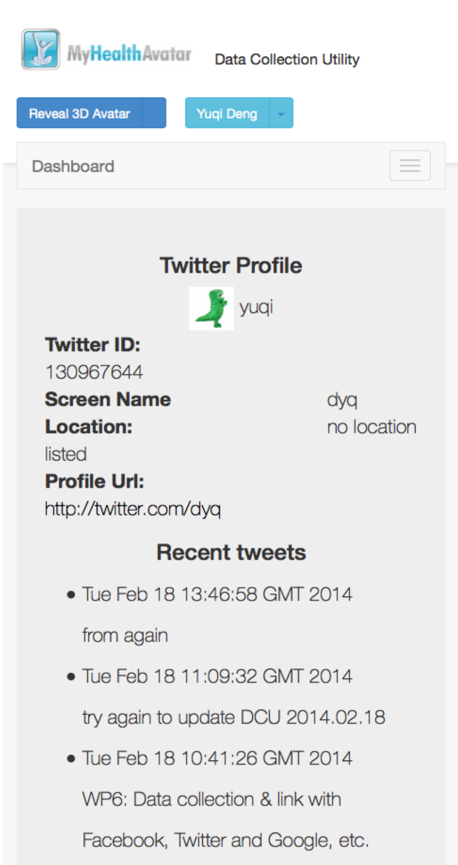

Figure 5.5 DCU UI from mobile browser

Three layers structure is used by DCU to separate the concerns and abstract business logic, data access and presentations. Following graph shows the major libraries used in each logical layer and the relationship between the layers.

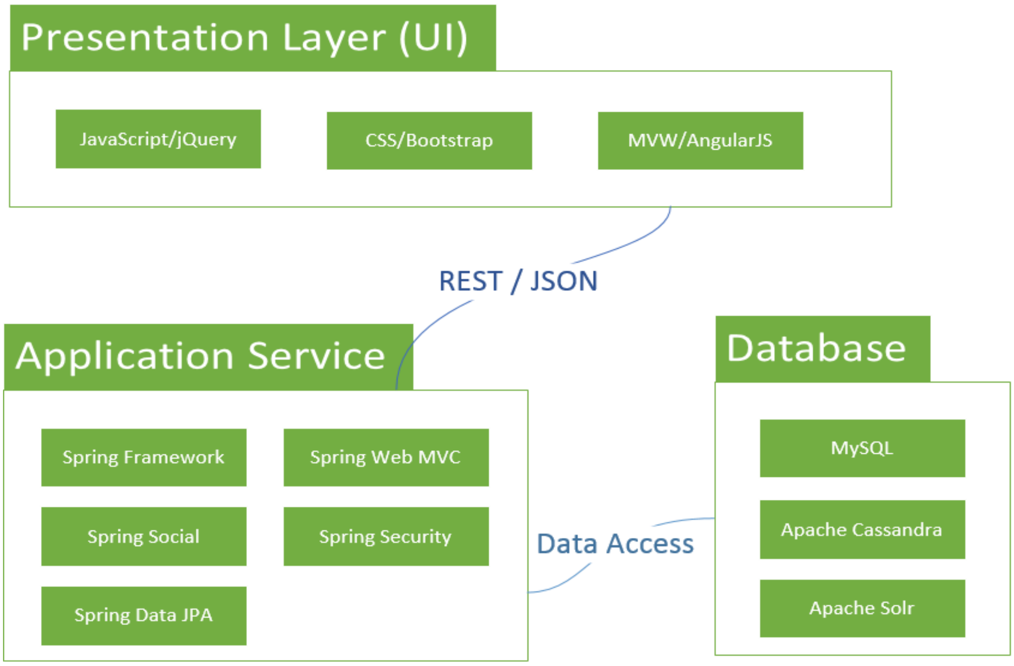

Figure 5.6 Three Layers Structure of DCU Implementation

5.6 Frontend Web UI

A set of tools are used to make the front end of MHA DCU interface, which are:

- **yo**: a scaffolding tool that generates the framework-specific scaffolds, which standardize and ease the project setup process.
- **grunt**: build, preview and testing tool which is similar to ant/maven in Java ecosystem.
- **bower**: dependency management for front-end libraries.

For frontend interface, libraries are used to avoid reinvent the wheel, they are state of the art toolset, which includes:

- **jQuery**: a arguably most popular front end JavaScript library which unify different browser API and gives developer a much easier and happier development experience.
- **AngularJS**: a JavaScript framework for developing dynamic web applications, it provide high level features like data-binding and dependency injection (DI).
- **Twitter Bootstrap**: a frontend CSS framework which lays the basis of MHA web UI and gives a modern flat look and feel.
- **Jasmine**: a test framework to test the frontend code and web UI behaviour automatically.

A set of domain specific programming language are used to generate targeting source:

- **SASS**:
- **CoffeeScript**:

In our UI application, we have a following folder structure:

- **app/**: a root directory for the MHA application
    - **index.html**: the initial base html file for MHA Angular application
    - **404.html**: the common resource/URL not found page to present to user
    - **favicon.ico**: the icon to be display in browser tab / title bar for MHA application
    - **robots.txt**: instructions for search engine to index MHA pages
    - **bower_components/**: directory for JavaScript/web dependencies, installed by Bower
    - **scripts/**: folder for MHA JavaScript/CoffeeScript files
        - **app.js**: main application code
        - **controllers/**: Angular controllers code
    - **styles/**: folder for CSS/SASS files
    - **views/**: a place for MHA Angular templates
- **Gruntfile.js, package.json, and node_modules**: configuration and dependencies required by the Grunt tasks
- **test/ and karma.conf.js/karma-e2e.conf.js**: a test runner and the unit tests for the project, including boilerplate tests for MHA controllers.

The source codes are tested using Karma and Jasmine. Karma is a good JavaScript test runner that is test framework agnostic, and DCU uses Jasmine as the testing framework. Don't Repeat Yourself (DRY) principle is applied, SASS and CoffeeScript are used when appropriate to make modelled front-end sources, which then give long-term benefit on readability and maintainability. After the entire test set passed, the front-end source will be built, minified and optimized for web which saves bandwidth and gives user a smooth experience.

5.7 DCU Data Post Processing Architecture

Data collected by DCU can be roughly categories as plain text, semi-structured data and structured data. Data need to be filtered and transformed to standard format in order to be stored, indexed and for future analysis and visualization.

DCU contains logic which does basic filter and transform of the collected data, following graph shows an abstract view of how DCU process these different data.

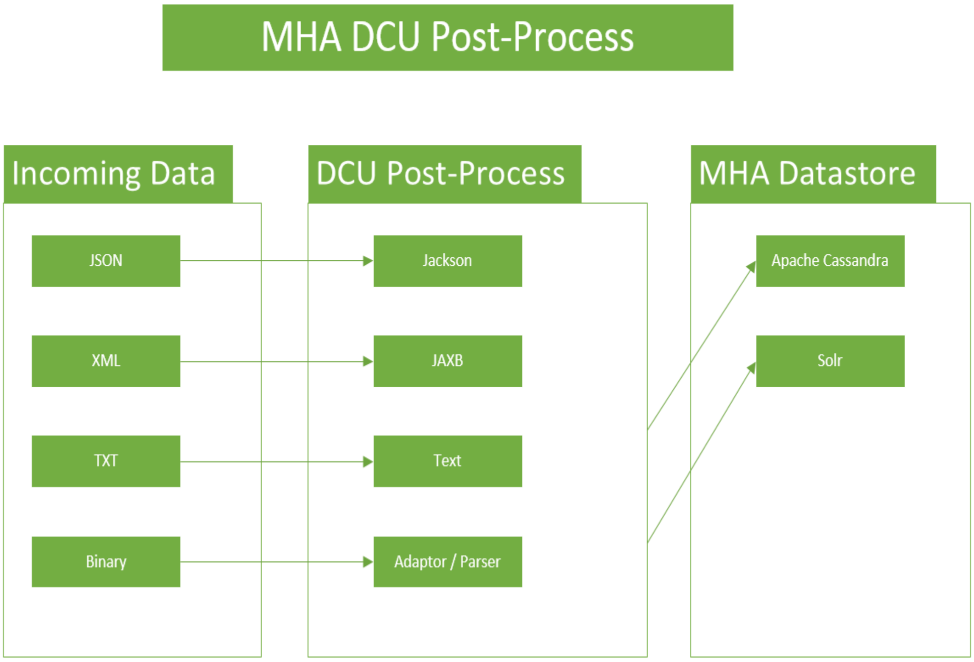

Figure 5.7 Post-process of DCU Implementation

For Rest services, normally JSON data format is available to collect from service providers. Jackson library is used to marshal between JSON strings and Java objects, it is very flexible and efficient in converting the JSON strings. Sometimes JSON is not available and XML is served by provider, DCU utilize Java Architecture for XML Binding (JAXB) to marshal between XMLs and Java objects. For plain text content, DCU itself pass them to MHA main system data mining logic to do ontology tagging and analyse. Depends on the format of binary data, it is normally stored in original raw format, with some of them also parsed and stored in other more meaningful format (e.g. XML, JPEG) when possible.

6 Conclusion

6.1 Summary of current Stage

The initial data collection utility web application is implemented and smoke tested, which will be generally available to public soon. The current architect and server configuration will be able to support initial expected batch of testing users, and expected to be scale according to the server load. The data collection utility is based on the technology stack used by MyHealthAvatar itself, and integration work goes on smoothly without any major challenges, which however will still take a bit of time to fully integrate.

The utility implemented as designed and fulfil the functional requirement, also internal testers are having positive experience using the flexible UI. As it collects to a group of initial chosen service providers, which tested to be working very well, however due to the testing time and resource limit, there is expected to be rough edges which will be fixed based on further testing phase.

6.2 Challenges

There are some typical challenging issues in implementing data collection solution for MHA platform.

Firstly, it requires more stable third party API to support the further implementation. As a lot of the APIs are either in beta version or in rapid development, it creates challenges for DCU to catch up with the provider's changes. The service providers normally impose some data access rate limitations on consumer applications, this potentially harm the DCU's ability to retrieve the user's data in a timely manner.

Various data source with emerging new technologies/devices and their new APIs, there are programming work to make new plugin/adapters for them. DCU is designed in a flexible way to cope with future changes; however additional coding and testing is still required to make DCU work with new type of devices/APIs.

With the number of MHA users growing over time, the scalability of DCU would be something need further refine. Although DCU has taken into account the growth of users and their data, it could still be unpredicted issues e.g. computational power and storage spaces of MHA.

DCU is only tested internally and user experience is not tested around different groups with different ICT skills, the design might need to be further improved based on future users' feedbacks. The new users will also help testing the DCU thoroughly to help report any bugs that might exist in the system.

Secondly, the inaccuracy and uncertainty of health data from diverse devices requires some advanced data aggregation algorithms to improve it. However, advance data aggregation algorithms are normally quite time-consuming, which may be suitable for mobile device uses.

Finally, privacy and ethical considerations are still a big challenge for public to free use MHA platform. This would limit the utilization of social networking resources.

6.3 Future Work

The primary future work would focus on the exploration of research value from data collection utitilities in MHA platform. Some questions like how to match uncertain raw health data to high level physical activity patterns, how to provide a sustainable and high-quality service with MHA platform, and how to optimize resource and QoS guarantee for the delivery of life-logging data streams are worthy to investigate.

Secondly, the MHA based mobile app is desirable to be implemented. This application should be with better user experience on interface, usability and effortless. Also, it should provide the function to easy share and join, reward users to sharing. The compatibility and extensibility of MHA apps are also considerable.

Finally, wearable computing technology is a fast-growing market. It is necessary to keep tracking other possible life-logging health data, such as Google Glass and Voice Input. The integration of these new sensors would be very benefit to MHA platform.

7 References

[1] S. Barnason, L. Zimmerman, J. Nieveen, P. Schulz, C. Miller, M. Hertzog, C. Tu. Influence of a symptom management telehealth intervention on older adults' early recovery outcomes after coronary artery bypass surgery. *Heart Lung*, 38, 364-376, 2009

[2] C.-C. Yang and Y.-L. Hsu. A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring. *Sensors*, 10(8), 7772-7788, 2010

[3] K. Y. Chen, D. R. Bassett, Jr. The technology of accelerometry-based activity monitors: current and future. *Med. Sci. Sports. Exerc*, 37, 490-500, 2005

[4] Fitbit. http://www.Fitbit.com (accessed Feb 6, 2014)

[5] Nike+ Fuelband. http://www.nike.com/us/en_us/c/nikeplus-fuelband (accessed Feb 4, 2014)

[6] Nike+ Sportsband. http://nikeplus.nike.com/plus/products/sport_band/ (accessed Feb 6, 2014)

[7] JawBone Up. https://jawbone.com/up (accessed Jan 6, 2014)

[8] F. F. Guo, Y. Li, M. S. Kankanhalli and M. S. Brown, An evaluation of wearable activity monitoring devices. Proceedings of the 1st ACM international workshop on personal data meets distributed multimedia, 2013 Oct, pp 31-34.

[9] Sokolova M, Jafer Y, Schramm D, Text mining for personal health information on Twitter. IEEE Second International Conference on Healthcare Informatics, Imaging and System Biology. 2012 Sep;pp112-118.

[10] J. Xiang, A. C. Soon, J. Geller, Monitoring Public Health Concerns Using Twitter Sentiment Classification. IEEE 2013 International Conference on Healthcare Informatics. 2013 Sep;pp335-344.

[11] H.H. Hsu, Y. F. Chen, C.Y. Lin, C. W. Hsieh, T. K. Shih Emotion Care Services with Facebook Wall Messages. 2012 26th International Conference on Advanced Information Networking and Applications Workshops, 2012 Mar;pp875-880.

[12] HealthVault. https://jawbone.com/up (accessed Jan 6, 2014)