Requirements Specification

WearWare Study Manager

Sponsored by Dr. Kyle Winfree and Dr. Eck Doerry
Mentored by Han Peng

Members:
Anton Freeman, Halcyon de la Rosa, Karina Anaya,
Keil Hubbard, Noah Olono

Version 1.0
11/22/2021

Accepted as baseline requirements for the project:

Client signature: _____ *Date:* _____

Team signature: _____ *Date:* _____

Table of Contents

1. Introduction

Heart disease is a major problem for many people in America. In fact, according to the CDC, nearly 1 in 4 people in America die of heart disease every year. In order to analyze the issue and prevent further deaths, researchers need to gather information on people with heart disease so they can figure out what best helps to fight it. This information gathering is often done using studies in which researchers employ the use of bioinformatics.

Bioinformatics is the field of collecting and analyzing biological data, which includes the use of heart rate monitors, fitness trackers, and other tools to gather data from study participants. The field of bioinformatics is incredibly important to anyone in the world as the results that it can produce can help to save lives. For example, bioinformatics could be used to find an exercise routine or figure out a medication which could help people recover from heart disease and thus help to save the lives of the previously mentioned 1 in 4 people who die from it. In fact, results from studies using bioinformatics could end up influencing the National Institute of Health's views, policies, and recommended treatments, meaning that they could affect the lives of every single person in America. Due to the importance of bioinformatics, the field has thousands of researchers, as well as two major annual conferences and two research journals dedicated entirely to it.

Our clients Dr. Kyle Nathan Winfree and Dr. Eck Doerry, two professors from Northern Arizona University, are working with the Wearable Informatics Lab at Northern Arizona University to try and introduce a new way of collecting data to the field of bioinformatics by creating a technology that will allow researchers to use Fitbits to collect information on study participants. Fitbits are small wrist-mounted devices that can be used to collect information such as heart rate, exercise, steps, and sleep data from people wearing them. Many of the current tools used in bioinformatics, such as heart rate monitors and fitness trackers, are expensive to use and can result in biased data; our clients report that often study participants are only given these tools for about a week, during which time the participants will alter their behavior due to unfamiliarity with these devices and the feeling of being observed. This results in inaccurate data when compared to a participant's normal life and behavior. Using Fitbits would lower the cost and amount of time that participants could wear Fitbits compared to other measurement devices would make it easier for participants to adjust to their vitals being measured, allowing researchers to get accurate results. Data collection with Fitbits could help almost any researcher that needs to gather heart rate, sleep, or fitness data, and as a result this project could produce value by allowing these researchers to get good results and help save lives. Additionally, the size of this business is very large, as our client believes that it could help anyone in the field who has an interest in collecting this data.

2. Problem Statement

Despite the amount of promise that our client's idea has, its current implementation has a number of issues that prevent it from fully reaching its full potential. In order to describe these issues, we will first describe our client's existing system's workflow as shown in Figure 2.1.

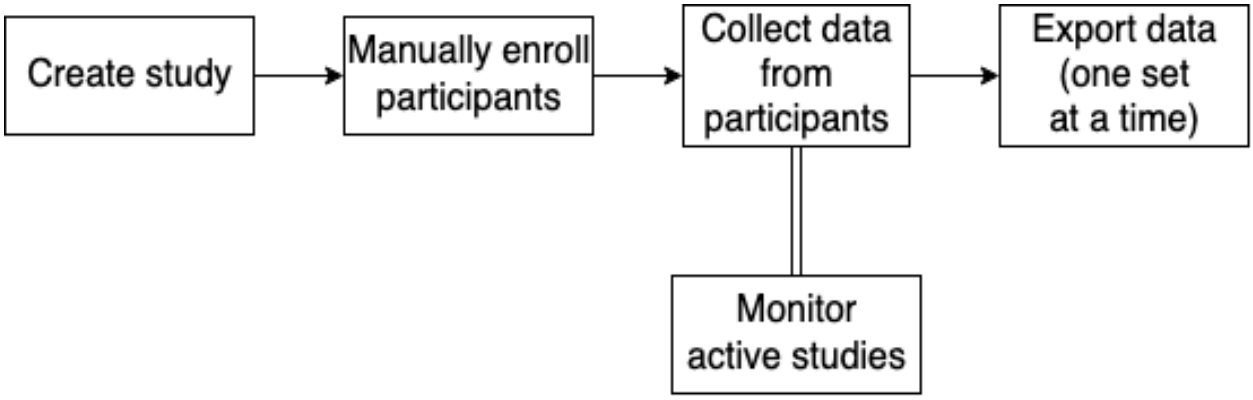

Figure 2.1: A diagram showing the current system's workflow.

Currently, our client's system allows researchers to create studies on a website for data collection and monitoring. Once these studies have been created, users can then manually enroll participants in these studies. After participant enrollment the researchers can then begin data collection over a specified time period, during which they can monitor the active studies in order to ensure everything is going smoothly and data collection does not have any issues. Finally, users can then export data from the studies in a CSV format to use for their own studies.

At first glance, this workflow seems effective. However, our client has outlined a number of significant issues with this system that result in it being at best inefficient and at worst inaccurate. These issues are:

- Export times are very long, which is expected but comes with several issues.
    - Long export times cause the system to time out, resulting in missing data sets.
    - The system also completely locks up while data is being exported, preventing users from doing anything else until the long exports are over. This also results in researchers only being able to export one data set at a time.
- Participant enrollment must be done manually by researchers, and participants cannot self-enroll.
- There is no way to communicate with participants using the system; all communication must be done outside of the system.
- Fitbit sleep data is not collected.

3. Solution Vision

Because of the issues with our client's current technology, we have been requested to replace the existing system with an entirely new, improved one. As a result, we plan to build a secure and reactive web app that will allow researchers to collect Fitbit data in the same way as before, but with many improvements in order to fix all of our client's current issues. Figure 3.1 presents a workflow diagram that shows how this new system will work.

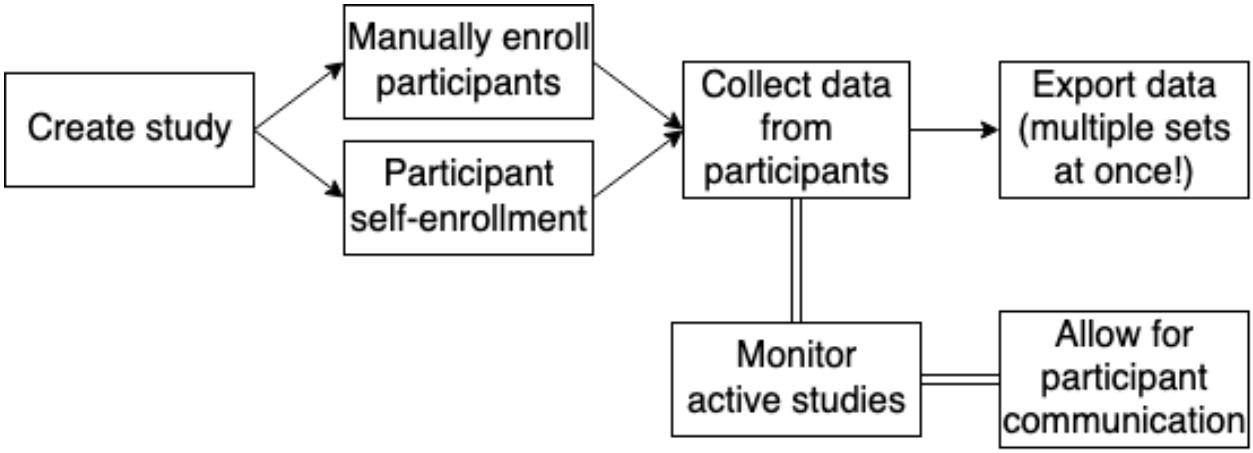

Figure 3.1: A diagram showing the new system's workflow.

As illustrated by this diagram, the workflow will be mostly similar to the previous system, with researchers still able to create studies, enroll participants, collect data during studies, monitor active studies, and finally export data. However, there are several significant improvements that we will make when compared to the previous system.

- In addition to manual enrollment by researchers, participants will be able to self-enroll via links.
- Our web app will allow researchers to contact participants via SMS message.
- Export times will no longer time out, so researchers will be able to get full data sets without anything missing.
- Exporting data will be done as a background task, so the system will still be available during data export.
    - This will also allow researchers to export multiple datasets at once.
- Fitbit sleep data will be collected.
- Researchers will be able to export data both on the website and via an API.

We have spoken to our client about these changes to the old system and have received approval on these improvements. As a result, we feel fully confident in our ability to create a quality project that will fit all of our client's needs moving forward.

4. Project Requirements

Now that we know what our client wants and what features we plan to implement, we need to figure out the specifics of what the features we are implementing will entail. In order to figure out what it is we need, we have held weekly meetings with our client to figure out these specific requirements that this project will have. We have since finished our requirement acquisition, and will list them all in this section. We will first begin with the domain-level requirements, which are very broad requirements that act more on the user level. From there, we will go into greater detail with our functional, nonfunctional, and environmental requirements.

Our domain-level requirements are split into two different types: website-level and API-level requirements.
Website domain-level requirements:

- Users will be able to create and manage accounts, studies, and enroll participants, including adding additional researchers to a given study.
- Users will be able to request specific study data to be exported in a predetermined format.
- The website will update study participant data using the Fitbit web API in the background until study conclusion or access is revoked.
- Users will be able to contact study participants

API domain-level requirements:

- Users will be able to authenticate using temporary security tokens
- Users can request data from their existing studies

4.1 Functional Requirements

4.1.1 Account creation/login/authentication/authorization

Users will be able to create new accounts, and login to those accounts to access the functionality of our web app and systems. To secure our systems and data we will implement user authentication for account access and user authorization for data access.

4.1.1.1 Account creation

Account creation will happen using OAuth2.0, with all usernames and passwords stored outside of our database.

4.1.1.2 User login

User login and authentication will also be handled through OAuth 2.0 which upon completion will return the user's token. This will then be used to retrieve the linked account from our system.

4.1.1.3 User authorization

User authorization will be handled by our local system by checking that the given user ID has sufficient privileges to access the given data.

4.1.2 Study creation and management

Users will be able to create studies using the web app and define the parameters of the study. Additionally, users will be able to modify studies they have ownership of or modification privileges for.

4.1.2.1 Study creation

Users will be able to create studies associated with their accounts and define the parameters for the study such as name, date format, number of participants, researchers and assistants associated with the study, and participant identification scheme.

4.1.2.2 Study modification

Users will be able to change or modify parameters of the study even after creation. These parameters will be the same as those defined in the study creation section 4.1.2.1. As an additional part of this requirement all modifications must be requested using our web interface.

4.1.2.3 Complete study export

Users will be able to export all data associated with a given study created and owned by them into a zip file. The structure of this zip file will be the same as that defined in section 4.1.5.

4.1.2.4 Study monitoring

Users will be able to monitor their studies, including viewing both the participant list and a GUI that shows participant activity.

4.1.2.4.1 Participant list

The study's participant list will contain a subject's identifier, most recent Fitbit sync time, amount of Fitbit syncs, and frequency of Fitbit syncs. Additionally, participants will be highlighted in yellow if they have not synced in 3 days and red if they have not synced in 1 week respectively. Participants will be selectable to show up on the GUI.

4.1.2.4.2 Activity GUI

The study's GUI will show a chart of the recent activity levels and sync times of participants which have been selected in the participant list. There will also be an option to show the activity and sync times of all participants.

4.1.3 Participant enrollment

Users will be able to add participants to their studies in one of two ways, individual or bulk upload. After uploading the information about the participant our system will generate a link that will be sent to the participant which will allow them to authorize our system to collect their Fitbit data

4.1.3.1 Individual participant upload

Users can add participants manually by filling out the individual enrollment information form on a given study. At a minimum this data will include the participant's name, age, and email address. With this information, our app will generate a link for the participant to use which will prompt them to confirm their information and authorize the data access and collection by the researcher which will also include information about the study and what data is being collected.

4.1.3.2 Bulk participant upload

Users can instead opt to bulk upload participants by selecting the bulk upload function and submitting a csv file containing the same information required for individual participant enrollment as defined in requirements 4.1.3.1 where each participant is a row and the columns are the required information. Our system will parse the rows and create enrollment links for each participant in the file. Once the participant has enrolled their participant account will be mapped to an incremented id in the desired study.

4.1.3.3 Participant self-enrollment

Users can create a link for participants to enroll, in which they will enter in their name, age, and email address. They will then be sent an enrollment link as specified in Section 4.1.3.1.

4.1.4 Requesting Fitbit data

An essential requirement of our system is to request participant Fitbit data using the official Fitbit web API.

4.1.4.1 Request cycle

Our application will attempt to pull Fitbit data no less than 1 time per 7 days and a maximum of 150 times per single hour as restricted by the terms of the official Fitbit API request limit per user per hour and data history specifications.

4.1.4.2 Requested data

Our application will request the following timestamped data for each participant: calories burned, distance traveled, Fitbit level, heart rate, metabolic units, number of steps, and sleep activity data. All responses will be recorded and saved to our database in accordance with the requirements defined in section 4.1.4 database requirements.

4.1.4.3 Request exclusion

Data from participants who have requested to leave a study will not be deleted; however, all future requests for Fitbit data from their account will be canceled.

4.1.5 Database requirements

4.1.5.1 Data export

Users will be able to export data from studies they have the proper access to. This section will outline the process for requesting and retrieving that data as well as the resulting format.

4.1.5.1.1 Exported CSV folder/file structure

The folder structure of a csv export will be as follows. Top level folder with name defined as the name-of-study_export-data.zip. Nested inside this will be folders named 'Fitbit_calories', 'Fitbit_distance', 'Fitbit_fb_level', 'Fitbit_heart', 'Fitbit_mets', 'Fitbit_steps', and ‘Fitbit_sleep_data”. Each of these subfolders will contain a csv file for each participant with the following naming scheme. Study-name_participant-id_start-date_end-date.

4.1.5.1.2 Exported JSON structure

The folder structure of a JSON export will be as follows. The exported JSON data structure will have several sub-structures named 'Fitbit_calories', 'Fitbit_distance', 'Fitbit_fb_level', 'Fitbit_heart', 'Fitbit_mets', 'Fitbit_steps', and ‘Fitbit_sleep_data”. Each of these sub-structures will likewise contain sub-structures for

each participant that contain a participant's ID and all collected data points for that structure's type.

4.1.5.1.3 Export from web app

Users will be able to request specific data from their studies from our web app which will export all records associated with the study in csv format following the same folder structure laid out in our client meetings and in requirement section 4.1.5.1.

4.1.5.1.4 Export from API

Users will be able to request specific data from their studies from our API which will export all records associated with the study in JSON format as specified in section 4.1.5.2.

4.1.5.1.5 Store Previous Exports

Our system will store web app exports for easy access without having to re-create the files when they are requested for download again.

4.2 Performance (non-functional) Requirements

4.2.1 Export in less or more time than the current system

This system must not have export times exceeding the equivalent export times on our client's current system. The exact specifications being 24 hours for exports of 10,000,000 (ten million) lines of csv and

4.2.2 Exports avoid timeout errors

The system must not have timeout errors for large data sets.

4.2.3 Website and data are available to use during export process

During any requested export the system must remain running and available to any users who use it.

4.2.4 Errors are human readable

All errors and export hiccups will send a notification to the user which will be readable in English instead of giving the user a potentially difficult to read error code.

4.2.5 Export notifications

Users will receive a notification on our web app once an export is completed in the web app. This notification will be implemented as a pop-up that will be displayed

immediately on export completion if the user is currently logged onto our web app or on the next logon if the user is not currently using the app.

4.2.6 Web app is navigable from any point

Web app pages must be interconnected with each other following a readable tree structure that connects each page.

4.3 Environmental Requirements

4.3.1 Hosted on AWS

Our client has required that the server for our solution be provided by Amazon Web Services. Additionally, all databases and storage will be hosted there as well.

4.3.2 Uses Ubuntu 20.04 LTS

According to our client our AWS instance must be running the latest stable long term support Ubuntu release, which as of this revision is 20.04.

4.3.3 Delete previous exports

Previously stored exports should be deleted as needed to make space when storage is low, with the oldest stored exports deleted first.

5. Potential Risks

In this section, we focus on the potential risks that this project faces. A risk is described on the likelihood of events that can affect the overall success of the development progress for the project. In order to fully understand the risks, we must analyze and deliberate over the probability and the severity of these risks. After deliberating with the client and discussed as a team to come to the conclusion of these risks that will impact the project:

- Fitbit API Access Changes
- Fitbit Company Closure
- AWS Price Changes
- Database issues
- Security failure

5.1 Fitbit API Access Changes

This project relies entirely on Fitbit's API, so any major changes made to the API could result in issues to our project. This seems unlikely, but in order to lower its severity we will not have our frontend directly connected to the Fitbit API. This will prevent any API changes from heavily affecting the use of the frontend while we work to fix the backend to fit with any changes as necessary.

5.2 Fitbit Company Closure

This project relies on Fitbit's continued operation as a company; however, there is always the possibility of the company's permanent closure. Even though this seems unlikely, Fitbit's closure would have a very high severity due to its product being the backbone of this project. If this were to happen, we would have to look into other similar devices to use for data collection instead.

5.3 AWS Price Changes

Even if the possibility is low, Amazon may change the prices of their web hosting services, which could possibly result in this project's price of maintenance becoming too high to continue. However, the severity and likelihood of this risk are both low, as Amazon is known for keeping low prices in order to stay competitive, and if this does occur it may not affect us too much. If prices are raised prohibitively high, we could look into an alternative either on a different hosting website or on a server that we would maintain ourselves.

5.4 Database Issues

This project plans on having many users in the future, and with increased scalability comes the possibility of multiple users using the application in multiple sessions at the same time. This would require the database to make many data export and import requests at the same time, possibly slowing the database. The probability of this risk is very high and fully expected, but its severity should be low with good management and use of the tools provided to us with our database system.

5.5 Security Failure

Security is one of the points that we must keep in mind to protect our users, both researchers and participants. Much like many tech companies, we want to prevent security breaches as much as possible. Participant data must be protected, making the severity of security breaches very high. However, we have taken several precautions while building our solution, so the probability of security failure is low. Additionally, we have chosen not to collect names from our participants in order to keep their identities safe in the case of a security breach.

6. Project Plan

Now that we have laid out our requirements and figured out the risks we may face as well as how we plan to deal with them, we can present a schedule for our current implementation plans. . Figure 6.1 shows a Gantt chart of our schedule, which contains all of our current milestones that we have laid out for this project with the approval of our client.

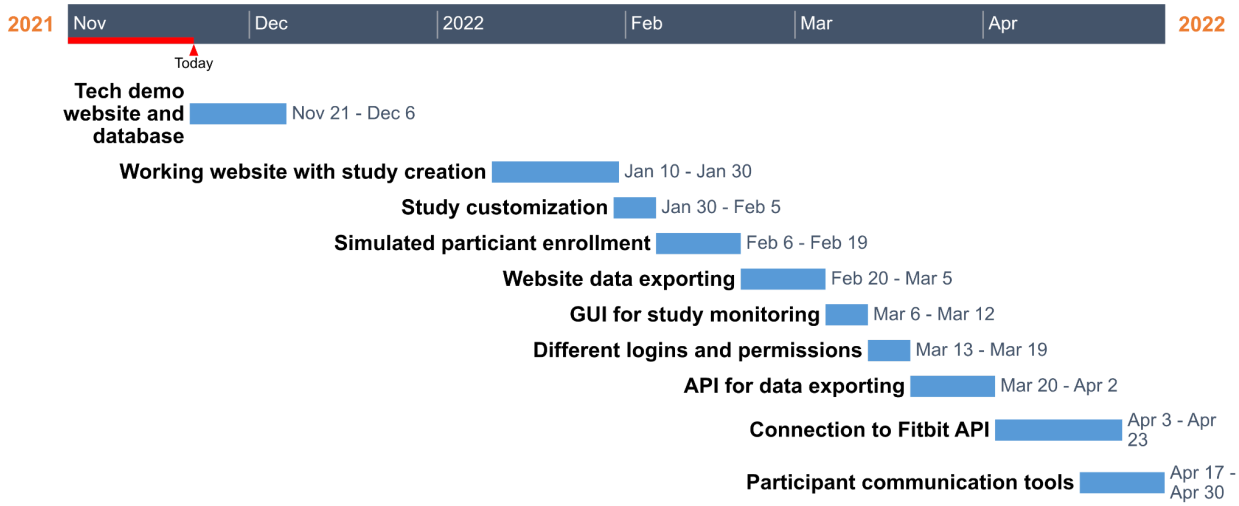

Figure 6.1: A Gantt chart showing our current implementation schedule and milestones.

Currently, we are working on creating a tech demo to show off how our website and database will work when fully implemented. From there, we will move on to using that tech demo as a foundation to create a fully functional website that allows users to create studies, followed by adding full customization for study creation as specified in Section 4.1.2.1. After that, we will allow for simulated participant enrollment, so that our client can try out enrolling participants in a study to ensure it works. Once that is completed, we will then allow for our web app to export data in a zip file using simulated data provided by our client.

Once we finally have basic data export done and thus have finished the foundation of this project, we can move on to the parts of our implementation that involve study management. This will begin with adding the study management tools as specified in Section 4.1.2.4. Afterwards, we will then fully add logins and permissions for researchers and assistants, as well as a participant login for self-enrollment and a way for researchers and assistants to enroll participants manually.

Then we will move on to our main stretch goals for the project as specified by our client. First, we will make an API that will allow for users to export data directly without using the website. After that, we will work on connecting our system to the Fitbit API so we can use data from there instead of only using example data provided by our client. Finally, we plan to create participant communication tools that will allow users to

communicate with participants via our web app. Once everything in this Gantt chart is completed, we will have a solid product that meets all of the requirements that we have listed in this document.

7. Conclusion

Bioinformatics is a very important field that could potentially save many lives from problems like heart disease. Our clients, Dr. Kyle Winfree and Dr. Eck Doerry, working with the Wearable Informatics Lab, wish to help improve this field by allowing researchers to use Fitbits to collect health data for studies. Our client's proposed use of Fitbit technology to collect information instead of using other more expensive tools like fitness trackers or heart monitors could result in researchers gathering more data than they normally would, while also allowing study participants time to adjust to having their health data collected and thus preventing biased data. However, our client's currently existing system has several issues that prevent it from being effectively utilized such as timeout errors, inefficient data exporting, and a lack of communication with participants.

To fix this issue, we will be building a web app that will replace our client's old system that will allow researchers to create studies, enroll participants in those studies, gather data from the participants over the course of a study, and then export that data for their own use. Some highlights of this solution that address issues in the previous system include the following:

- Participants will be able to self-enroll via links or email, while researchers will still be able to manually enroll them as well.
- Our web app will allow researchers to contact participants via SMS message.
- Export times will no longer time out, so researchers will be able to get full data sets without anything missing.
- Exporting data will be done as a background task, so the system will still be available during data export.
    - This will also allow researchers to export multiple datasets at once.
- Fitbit sleep data will be collected.
- Researchers will be able to export data both on the website and via an API.

At this current point in the project, we have finished all the preliminary research that we need to do in order to build this project, including figuring out what technologies we plan to use, gathering our requirements, figuring out what risks this project may have and how we intend to avoid them, and showing all of our progress to our client for his approval. Moving forward in our implementation plan, we plan to finish a tech demo that will show off both an example front-end for our web app and a database that will hold example data given to us by our client. After that, we will use that tech demo with feedback from our client to make a fully fledged web app following the proposed schedule from Figure 6.1. We believe that with this plan, we will be able to make a fully functional web app for our client that will fit all of his needs and requests.