**Air Quality Mapping**

Final Report for CS39440 Major Project

*Author*: Robert Mouncer ([rdm10@aber.ac.uk](mailto:rdm10@aber.ac.uk))

*Supervisor*: Dr. Neal Snooke (nns@aber.ac.uk)

24th April 2018

Version 1.0 (Draft)

This report is submitted as partial fulfilment of a Meng degree in  
Software Engineering (with integrated year in industry) (G601)

Department of Computer Science

Aberystwyth University

Aberystwyth

Ceredigion

SY23 3DB

Wales, UK

**Declaration of originality**

I confirm that:

* This submission is my own work, except where clearly indicated.
* I understand that there are severe penalties for Unacceptable Academic Practice, which can lead to loss of marks or even the withholding of a degree.
* I have read the regulations on Unacceptable Academic Practice from the University’s Academic Quality and Records Office (AQRO) and the relevant sections of the current Student Handbook of the Department of Computer Science.
* In submitting this work I understand and agree to abide by the University’s regulations governing these issues.

Name Robert Mouncer

Date 27/04/2018

**Acknowledgements**

I am grateful to Dr Neal Snooke, my project supervisor, for overseeing the project and guiding me in the right direction when I’ve gone off course.

I am grateful to Information Services and CSSupport for the hosting of data and websites for this project to be a success.

I’d like to thank Riversimple, my industrial placement of 2016 to 2017, for the idea of the air quality mapping project and allowing me to pursue it.

**Abstract**

Air pollution has had a large impact on the world from hazardous gases effecting the atmosphere to the death of millions of people each year. This has become an increasingly concerning problem in recent years as the effects have been researched and can cause damage to multiple entities. 40,000 deaths within the UK each year have been linked back to air quality levels [XXX] with a large portion of deaths happening in major cities such as London. Janurary-March 2017 it was estimated that nearly 40 million vehicles are on Great British roads [XXX]. This has a large involvement on the air pollution levels within the UK, but these vehicles may help provide a solution to this problem.

The Royal College of Physicians released 14 steps needed to improve pollution levels, one of these was to “monitor air pollution effectively” and to “educate the public” [XXX]. This projects purpose is to provide a means of educating the public of the air quality on public roads. The project will contain two systems, a monitoring system that will be used within vehicles to collect a range of data while travelling and a visualisation system to be used to educate the public in a proactive way. The monitoring system has been designed with the idea of the system being implemented on a small percentage on public roads to help build a dataset.

This report will explain the procedure taken to complete these two systems and the process leading up to their success.

Overall the project was a success with two working systems, though the overall functionality is limited the main goals of this project have been completed. This project provides a proof of concept to the idea originally from Riversimple.

**Contents**

1. Background, Analysis & Process 6

1.1. Background 6

1.2. Analysis 10

1.3. Process 11

2. Design 14

2.1. Overall Architecture 14

2.2. Hardware Design 15

2.3. Monitoring Software Design 17

2.3.1. Data Logging (Data\_logger.py) Design 19

2.3.2. 19

2.4. User Interface Design 19

2.5. Other Relevant Sections 19

3. Implementation 19

4. Testing 20

4.1. Overall Approach to Testing 20

4.2. Automated Testing 20

4.2.1. Unit Tests 20

4.2.2. User Interface Testing 20

4.2.3. Stress Testing 20

4.2.4. Other Types of Testing 20

4.3. Integration Testing 20

4.4. User Testing 20

5. Critical Evaluation 21

6. Appendices 21

A. Third-Party Code and Libraries 22

B. Ethics Submission 23

C. Code Samples 23

Annotated Bibliography 23

# Background, Analysis & Process

## Background

During the industrial placement at Riversimple from September 2016 to September 2017 a telemetry unit was being built to collect data from a hydrogen vehicle to monitor how customers were driving and report any fault with the vehicle. An employee at Riversimple mentioned an idea of being able to collect air quality data with the telemetry unit and use that data to show the public a comparison of air qualities around the country. With permission from Riversimple to pursue this project, the aim was to create a method of monitoring and visualisation air quality data inspired by the telemetry unit.

During the industrial placement it became clear of the state that the air pollution in the country was in due to personal and public transport. The motivation of this project stems from this and has developed over time. With little experience of such systems, the challenge was part of the motivation to complete this project and ensure its success.

**Background Preparation**

There was little background preparation for this project until the project had been reviewed and accepted as a major project. It began between the project being accepted and the start date of the major project.

The telemetry unit was built using a Raspberry Pi Model 3B+ (RPI). With little experience of using a RPI, preparation was carried out to ensure general knowledge of the microcontroller was known before work was carried out.

The operating system (OS) for the RPI were a vital part of the preparation, it needed to be decided what OS was to be used to suit the task at hand. Most RPI operating systems run using a variation of Linux. The time leading up to the project hand out, comparisons were made between the OS’s.

Supported protocols for the RPI were investigated, it turned out that most protocols could be used but had to be connected to the dedicated pin on the RPI board. If a protocol was not supported locally, a RPI HAT could be used to allow for support. A RPI HAT is an add-on board for the RPI, it stands for hardware attached on-top.

Hardware components needed to be investigated to ensure that it was possible to achieve what had been set out. Using the microcontroller, location and air quality data needed to be collected. Knowledge was known about GPS’s and being able to get the current location using them, but little was known about air quality sensors. Air quality sensors were found online but most of them needed additional components or were not suitable for the project. The accurate sensors and multi-gas sensors were very expensive and due to the extent of the project were not suitable. Several sensors were appropriate for the project. From this I knew I could create the hardware needed for the project, and if they were not very accurate or reliable, it would still prove as a proof of concept.

Having previous used a Google Maps API for commercial purposes, it was known the license agreement was not very permissive, research on other online map providers was conducted. OpenStreetMap was found to be very permissive and only required recognition on the webpage it was being used for.

Research was conducted on similar products to understand what solutions had been created and what to avoid for copyright purposes.

A blog was created to record what had been learnt and to keep track of any work that had been completed. This blog was hosted on the Aberystwyth University public\_html directory. My supervisor was given access to this to keep up to date with what had been worked on.

A GitHub account had already been created for personal use and this was used for the major project, though the repository was set to private.

**Similar Systems**

When researching the monitoring hardware and implementation it was discovered that a team measuring air pollution within London using pigeons [XXX], the pigeons would wear small backpacks with air quality sensors and a GPS. The type of hardware is the same, but the deployment is different.



Figure - Pigeon Air Patrol

Pigeon Air Patrol also has an interactive map which shows the air pollution across London, no values are shown, rather an indicator is used whether the area has “fresh air” or a certain level of pollution such as “moderate” or “high” as seen in Figure 1.

Plume labs, the same company that deployed Pigeon Air Patrol also are working on a device to measure air quality and location. It is a smart air quality tracker designed to be attached to a user’s possession, such as a bag or bicycle. This has not yet been released so information on it is limited.

The Department for Environment, Food and Rural affairs (DEFRA) have a pollution mapping website that is forecasted by the met office. The information states that the data is collected in various regions from monitoring sites and generated from current air quality issues.

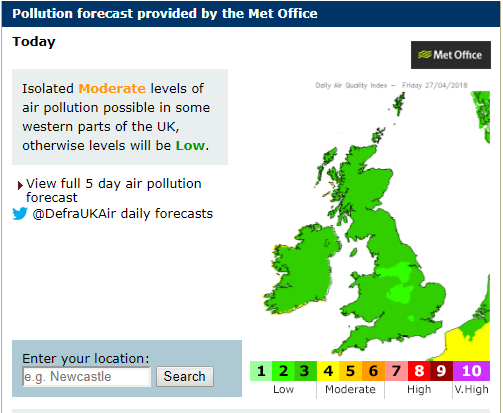


Figure - UK-air DEFRA screenshot

Figure 1 shows a forecast of the pollution over the UK, with a non-interactive map, the only interaction is a search functionality. Once a search has been complete, the map becomes interactive using google maps. This then shows values of pollution at different points, rather than the whole area, as can be seen in Figure 2.

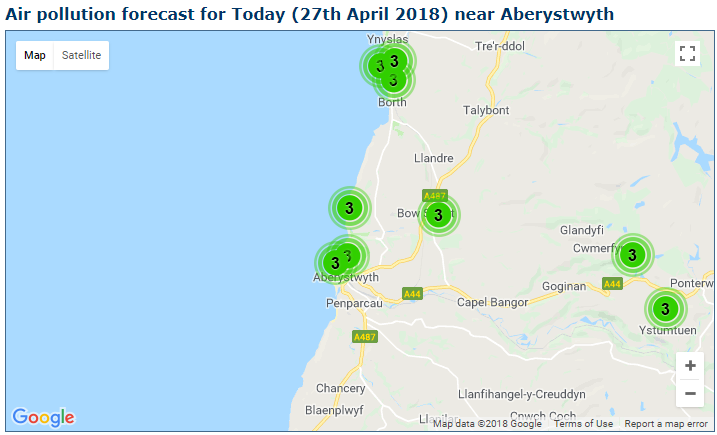


Figure - interactive map on UK-air DEFRA

The difference Figure 3 and the major project is that fixed monitoring stations weren’t used. The idea was to use vehicles as monitoring stations as even for local use, can still collect a lot of information regarding the air quality. The visualisation of the map was designed as a heatmap/contour map.

The DEFRA website also has links to Wales, Scotland and Northern Ireland air quality sites. The Welsh site lacks even more functionality than the UK site. The Welsh site uses less monitoring stations than Wales has on the UK site. The Scottish site has the most interactive welcoming screen.

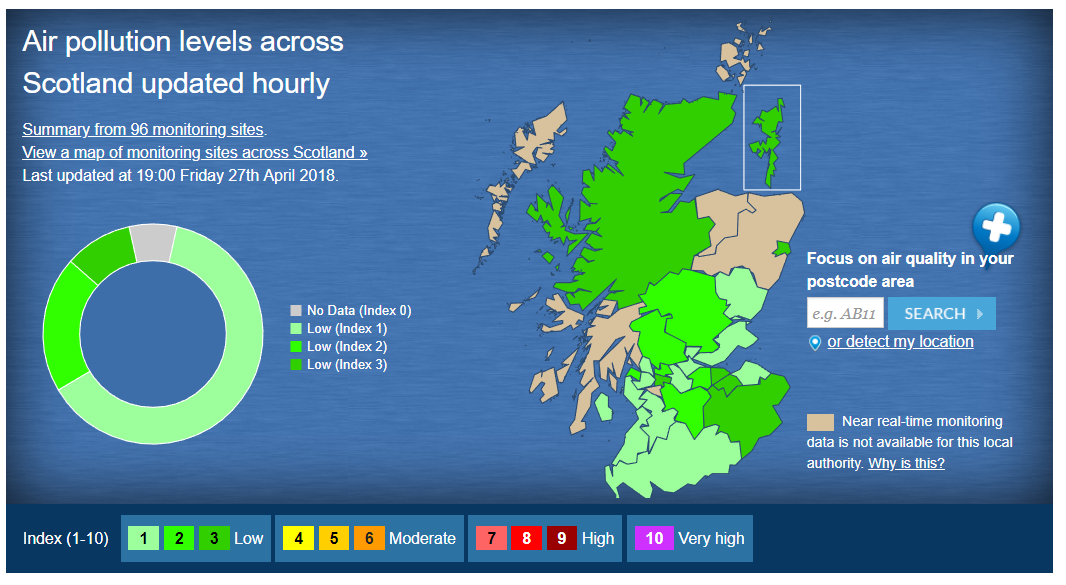


Figure - Interactive welcome screen Scottish Air Quality

Figure 4 allows users to select their location by clicking on their province. The design for the visualisation site is different to this and resembles no correlation. The interactive map is the same as the welsh and UK site, this is not what the major project aimed to achieve.

Riversimple contacted me before the major project started to make me aware of what they thought was a similar website. They had been in contact with the company at an event and were made aware of the website before I was.

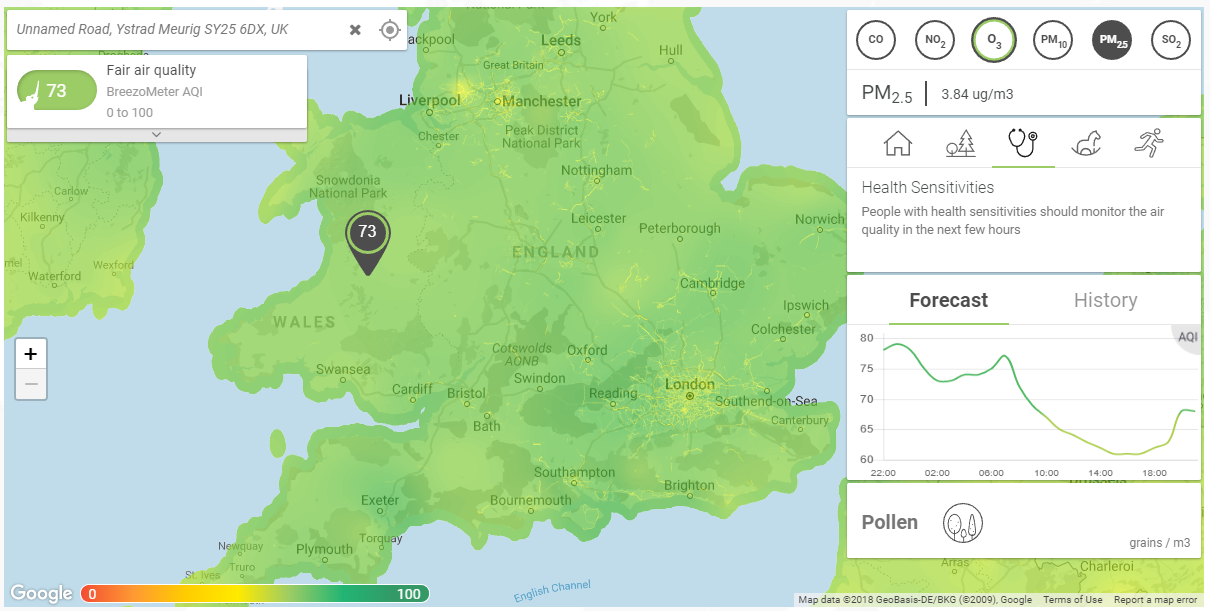


Figure - Breezometer interactive map

Breezometer uses weather/air pollution stations and machine learning algorithms to predict the values as seen in Figure 5. The monitoring system designed for this project is to use real world values rather than from monitoring stations and predicting the values. The interaction on Breezometer is very good and offers a lot of information. It shows different pollutants as one value on the map, but on the information menu, you can find individual information on the pollutants. Information is also given on the health issues and sensitives, this is a similar idea that was not fully implemented on the final version of the visualisation tool.

## Analysis

After studying the background research, it was drawn to a conclusion that this project would be possible within the timeframe. Various assumptions were made during the start of the project. The project would be split into two major components:

1. Monitoring system
2. Visualisation System

The monitoring system would collect the data and would need to be designed to be suitable within automobiles. The output of the monitoring system would be sets of data from journeys that would need to be uploaded to Aberystwyth University MySQL server.

The visualisation system would need to use the data collected from monitoring system that would be stored on the MySQL server. This would then be displayed to the user in a proactive way.

It was decided to use the Aberystwyth University MySQL server to store the data and public hosting for the website as access to it was very easy. It was already set up to suit the needs of this project.

Implementation of the project would require several different skills ranging from the use of hardware components in the creation of the monitoring system to web development.

The required skills needed and those to be developed are:

* Hardware selection
* Hardware creation
* Linux installation, command line and configurations
* Network Administration
* Python skills
* MySQL skills (Database Administration)
* Web development skills

The hardware would consist of a Raspberry Pi Model 3B+, air quality sensor and a GPS. The Raspberry Pi Model 3B+ was used because this resembles the microcontroller that Riversimple use for the telemetry unit that is currently being developed. This would be controlled by a Linux distribution called Raspbian Lite, meaning there is no GUI (Graphical User Interface) and would require an SSH (Secure Shell) connection or serial connection to be communicated with. Raspbian Lite was selected due to the RPI being a lightweight microcontroller, GUI is resource intensive and when the monitoring system is working a GUI is not required. Using the Lite version of Raspbian would require Linux command line skills, it was also chosen to improve these skills for future projects.

The RPI supports many different protocols for components such as I2C (Inter-integrated Circuit), SPI (Serial Peripheral Interface) and provides GPIO (General-Purpose Input/Output). This was another reason that the RPI was chosen. A Raspberry Pi HAT can be used if a hardware component was not supported by default such as CAN (Controller Area Network) which is most often used in the automotive sector.

As it wouldn’t be viable to create the software running on the RPI through the SSH, a samba server needed to be set up to allow a standard client-server interaction. The samba server would host the files while it was possible to connect to the server from another computer and edit the files using an IDE or advanced text editor.

It became clear that the software on the RPI would not need to be that advanced but would rely on the Linux distro to be correctly configured to allow the scripts to run on start-up and upload data collected. It was decided that only two files would need to be created to collect and upload the data.

The use of the Aberystwyth University resources would need to be considered. Hosting the webpage on the public\_html directory would be needed as it allows the webpage to be accessed from anywhere, this would be useful when giving demonstrations or presentations on the current state of the project. The hosting server also allows the use of php scripts, this would be used to get the data from the server and not allow the user access to the database, it will be handled server side.

It was decided that the visualisation application would use OpenStreetMap (OSM) as the map provider. The online community for developing OSM applications is quite large due to the data being geological data being open to anyone to use with only recognition needed to be given. The data provided by Google Maps is copyrighted by many organisations and it wasn’t clear whether I would be limited by Googles API or Terms.

The visualisation application would need to show the data in a proactive way for easy educational purposes. It was proposed that either a contour or heatmap would be used to display the data for easy comparisons of different areas. OSM supports a variety of plugins, including ones for creating heatmaps, this was another reason to choose OSM over Google Maps.

Security was an issue with this project at this point as connection to a personal database is needed on both systems being developed. At this point security solutions had not been considered.

The functionality of each system had been defined by this point, how the functionality would be implemented was not.

The objectives of work at this point in the project were:

* Select the appropriate hardware to work with the model 3B+ RPI
* Install Raspbian Lite onto an SD for the RPI
* Set up a Samba Server on the RPI for easy development
* Hardware components design and assembly
* Design of Monitoring system
* Development of the monitoring system
* Design of Visualisation system
* Development of the Visualisation system
* Testing of both systems

## Process

At the start of the project, the process instinctively took off with a waterfall approach. Rather than starting with the requirements, the project commenced in the design stage, without requirements being harnessed. This was a bad start to the project as requirements had not been set in place and therefore varied from one choice to another. About half way through the design phase, a new process was decided.

Having previous experience with agile methodologies within a team, it was known these methods work, it’s all about suiting the process to your needs. The difference being that this was a solo project, no matter the methodology chosen, it would need to be tailored to the project.

The problem that occurred were priorities were changing, a process was needed that would fixate on tasks and not change unless it was absolutely needed. The process that was chosen was Kanban. A document was created for the Kanban process to document how the methodology was adapted for a solo project. The approach created was to focus on software development as much as possible, and only focus on other tasks where necessary.

Trello is an online project management application that will be used for the digital boards and cards it can visualise. This online application will be used to manage the project, this includes harnessing requirements of work to be completed. One air quality mapping board will be used.

The requirements/tasks would be created as cards and these cards would move around the board dependent on what stage they’re in. The use of “stories” was not considered, but rather just a task on the cards. The stages were defined as columns, five columns were created:

1. Work – At the start of the project, this would contain all cards that would need to be completed. If a problem or task arose a new card would be created in the work section. The cards would move to the appropriate column when work had commenced on the task.
2. Work in Progress (WIP) – If a card was being worked on then the card would move to this column. A maximum number of 3 cards would be worked on at the same time, this was to prevent many tasks being allocated at once. If cards are dependent on each other and need to be worked on, then they would move to the “depends on another task column”.
3. On hold – If a task needs to be put on hold for any reason (prioritise work) then this column should be used. This column was a last resort and would only be needed when necessary.
4. Depends on another Task – If work had begun on a card and it was dependent on another task being complete first, then the card would move to this section. An example of this was developing python code to connect to an I2C bus but the I2C bus had not been connected.
5. Complete/Done – Once a task had been completed then it will move to this column.

A Gantt chart was created at the start of the project to measure progress, though the completed stories would show progress in its own way, it did not show effort complete and remaining effort. The Gantt chart was created for the first nine weeks of the project (Appendix A), once the first nine weeks had been completed then another up to date Gantt chart would be created for the remaining time (Appendix B). The Gantt chart predicted the time it would take to complete each major task. It would be often updated to show work completed on each task with the use of a percentage. The Gantt chart was a template provided by Microsoft Excel.

A version control system was used throughout the project, this was GitHub. All work regarding the project was to be stored and version controlled. This was to prevent any work to be lost or accidentally deleted. It is a way to mange work without having multiple versions scattered across multiple devices. In general, it makes the process of completing the project a lot easier.

# Design

## Overall Architecture

As stated in the analysis (Section 1.2) the overall system is split into two systems. The two systems do not communicate with each other but rather share the same data that is hosted on the Aberystwyth MySQL server. The two systems are the monitoring and visualisation systems, these could be split into further subsystems:

1. Monitoring System
   1. Hardware Architecture
   2. Data Collection
   3. Data sharing
2. Visualisation system
   1. User interface
   2. Data retrieval

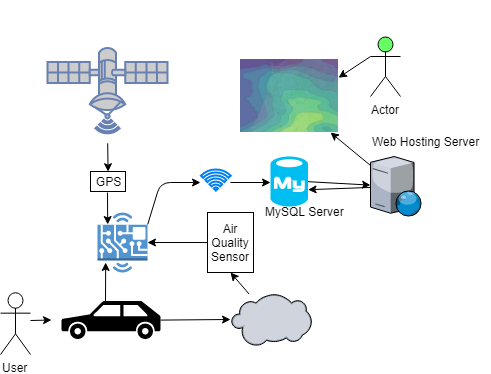


Figure – overall architecture of Air Quality Mapping

As Figure 6 shows the overall architecture of the project. The device is designed for vehicles, those vehicles will produce pollutants which the air quality sensor will detect and communicate a reading to the microcontroller (RPI). The GPS will do the same task when a location has been found. An NMEA sentence will be transmitted to the RPI and will need to be decoded. The air quality and location of the reading will need to be matched with one another, so the data isn’t rendered useless. Once a connection to an online network has been made, the data that has been captured should be upload from the RPI to the MySQL server. Once the data has been uploaded to the server it should be removed from the device. The data will remain on the server and be queried by the web visualisation application. The data will then be displayed to the user using OpenStreetMap and a heatmap overlay.

As the data is not very detailed and will only require several parameters, the MySQL table will not be very advanced.

Rather than a web application, a standard desktop application was considered with the use of java or python. This would not be viewable through a browser but rather its own application. This was dismissed when considering the customisability of using HTML (Hypertext Markup Language) and CSS (Cascading Style Sheets) on webpages.

Hosting the webpages and SQL server locally rather than the university network was considered but the two servers were already set up for needs of the project.

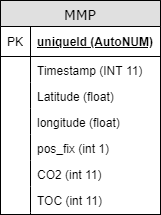


Figure - Table UML

As the MySQL server only needs to host a small amount of data, only one table would be necessary. Every table needs a unique id to identify tuples from one another. The attributes in Figure 7 are very minimalistic and does not include a lot of information that can be retrieved from the RPI, GPS and air quality sensor. Both the monitoring and visualisation systems will require access to the MySQL sever, so the data stored needs to fit the requirements of both.

## Hardware Design

Having little knowledge of hardware made the design process difficult. One of the only decisions made entering the design stage was a RPI were to be the microcontroller. The RPI could communicate with multiple protocols and had WiFi technology built into the board.

The process started by looking at the niche component being the air quality sensor. Many sensors were considered in this process, there were several reasons why some sensors were disregarded:

* Detecting if a gas was present rather than returning a concentration value
* Detecting gases that were no harmful to the environment of human health
* Too expensive
* Required the use of third party libraries or required RPI HATs to communicate with, essentially a non-supported protocol was being used.

The ams IAQ-CORE P sensor (Appendix C) suited the needs of the project using an I2C bus that was supported by the RPI. The sensor returned a reading of TVOC (Total Volatile Organic Compounds) which are produced from several sources and are hazardous to both the environment and human health. Along with the TVOC reading, a carbon dioxide estimation was made by the sensor and could be returned in the I2C message. This was the sensor that was implemented in the design and implementation.

A GPS was needed, this was an easy decision as the GPS modules would do the same task. The deciding factors were to find a GPS module that had a supported protocol on the RPi and that would be easy to wire on a breadboard. The PmodGPS receiver (Appendix D) was chosen as it fit these requirements. It uses the UART protocol which is supported by the RPI and is a variation of a serial connection.

Once the components were selected, a diagram was created using Fritzing, it is an open source computer aided design software package that allows the design of circuits.

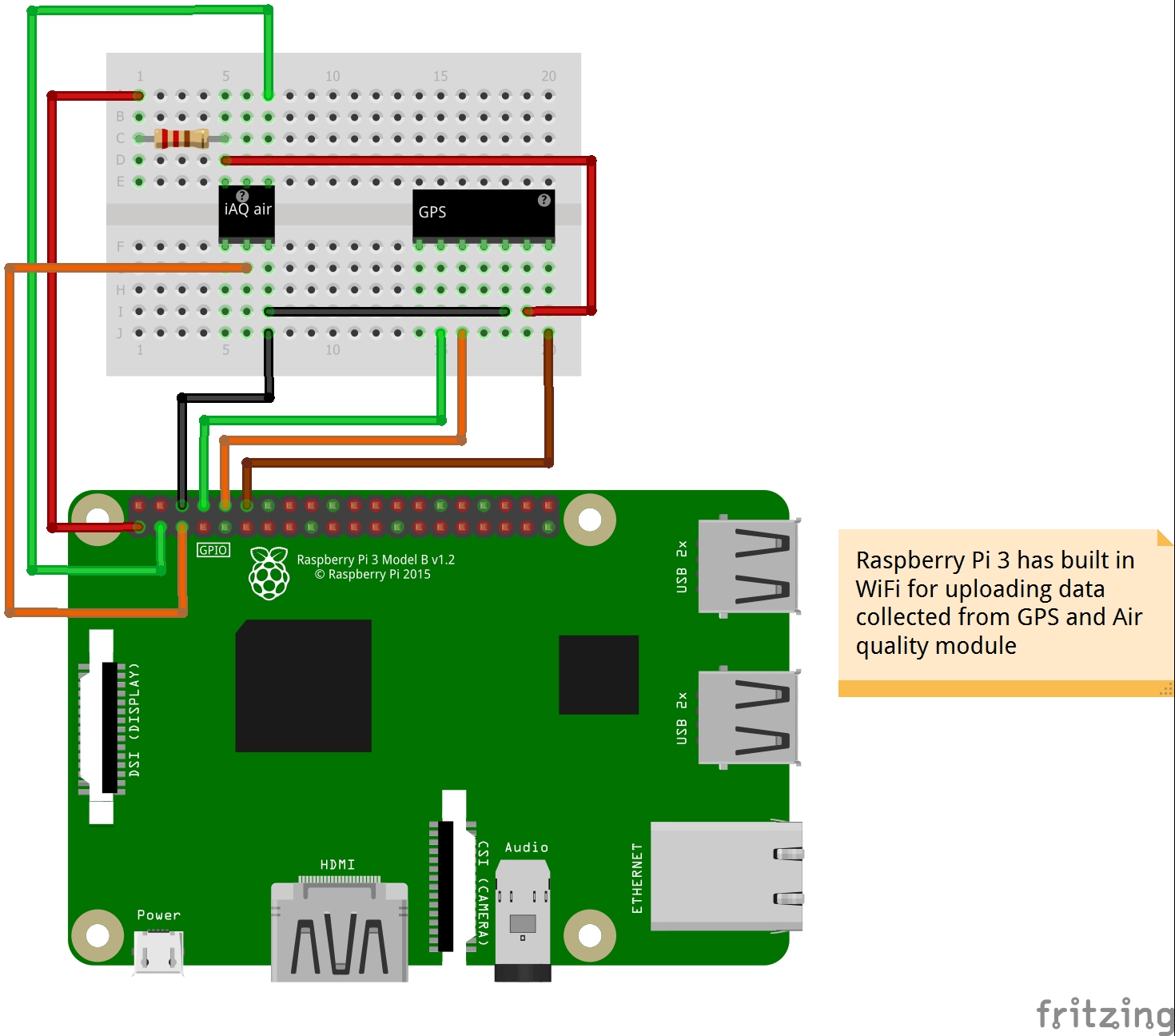


Figure - First design of Hardware

|  |  |  |
| --- | --- | --- |
| RPI Pin | GPIO number | Use |
| 1 | - | 3.3v power used for the Air Quality Sensor and GPS. |
| 3 | 2 | I2C data bus connected to the Air Quality Sensor. |
| 5 | 3 | I2C clock connected to the Air Quality Sensor. |
| 6 | - | GND used for 0v for both components. |
| 8 | UART0 TX | The transmit serial line connected to the receive line on the GPS. |
| 10 | UART0 RX | The recieve serial line connected to the transmit line on the GPS. |
| 12 | 18 | General output connected to the reset pin on the GPS. |

Figure - RPI Pinouts

Figure 8 shows the first design that was created for the hardware. It consists of the use of the serial and I2C dedicated pins. Figure 9 details the use of each pin on the RPi and it’s use. This was the design when development first began bus it was soon realised that this would not work and the design for the monitoring system hardware would be changed.

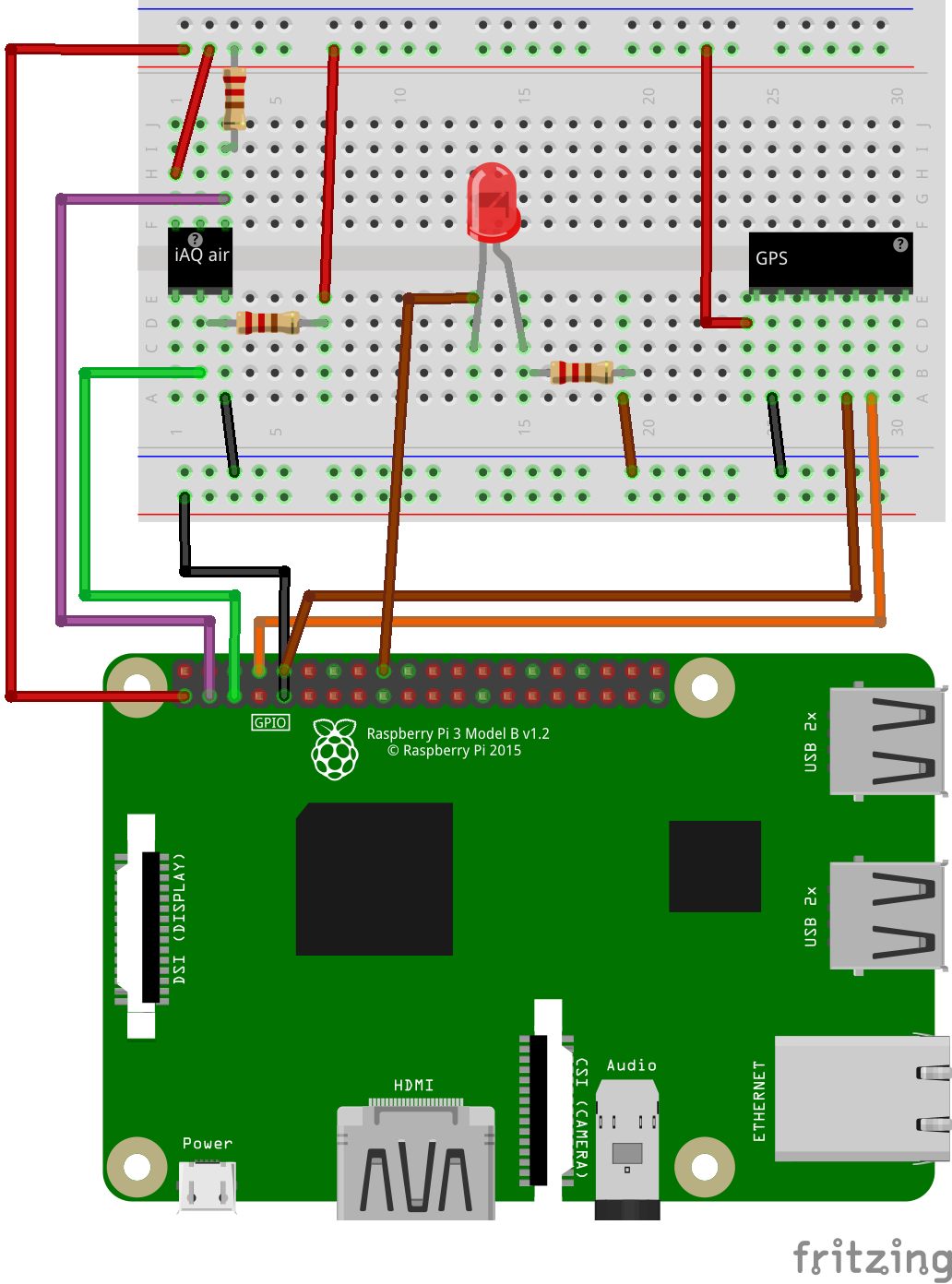


Figure - Final design for the Hardware

|  |  |  |
| --- | --- | --- |
| RPI Pin | GPIO number | Use |
| 1 | - | 3.3v power used for the Air Quality Sensor and GPS. |
| 3 | 2 | I2C data bus connected to the Air Quality Sensor. |
| 5 | 3 | I2C clock connected to the Air Quality Sensor. |
| 8 | UART0 TX | The transmit serial line connected to the receive line on the GPS. |
| 9 | - | GND used for 0v for both components. |
| 10 | UART0 RX | The recieve serial line connected to the transmit line on the GPS. |
| 18 | 24 | General output used to flash LED, this will show what state the data logger is currently in. |

Figure – Final RPI Pinout

Figure 10 shows the final design for the hardware. The changes made to the I2C were to include a pull-up resistor on the clock and data bus. This is necessary for I2C as the interface can pull the signal low but can not drive it high, the pull-up resistors are used to restore the signal to high when there is no low signal.

The reset pin on the GPS was disconnected as it was never used within the implementation. The GPIO output has been moved to pin 18 and will control an LED to signal the user what state the data logger is currently in.

## Monitoring Software Design

The software for monitoring is split into two parts. The data logging which consists of retrieving data from the GPS and air quality sensor then writing it to file and the uploading of data collected by the logger to the MySQL server.

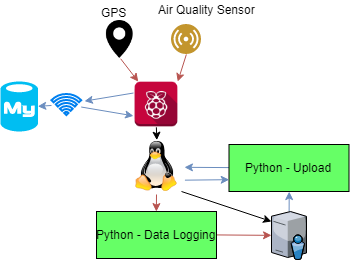


Figure - RPI Overall Software Design

The language chosen to develop the software for the monitoring system was python, this was due to its vastly available libraries, functionality with hardware and online support. As the operating system on the RPI will be a distribution of Linux, python could be easily installed. The python files would be created in a directory of the default user with a secure password, the default user is Pi. The style of programming would be procedurally, python allows multiple programming styles, but procedurally programming favours iteration, which both files were predicted to use most of the time.

As shown in Figure 11 the data logger python script would take readings from the GPS and air quality sensor. These readings would be passed to the OS from the RPI hardware. This data would then be stored onto the local directory of the Pi user. The upload file will be run on occasion to check if a connection has been successful to the MySQL server. If a connection to the server is a success, then uploading should begin. Once a file has been uploaded to the server then it should be removed from the RPI.

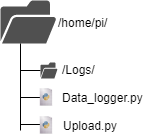


Figure - pi user file structure

The file structure for the pi user is shown in Figure 12, this contains the two files needed to run the monitoring system. The files are shown to be in the pi users directory. The logs folder has been designated for log files for the data logger to store files.

The files stored in the logs directory were designed to use the same attributes as the table attributes shown in Figure 7. On the first lone of the log file the attributes were to be listed. The data would then follow in the same order as the attributes on a separate new line each time data was appended to the file. An example of the design of the log file is shown in Appendix E.

The log file name was designed to be unique, it was to be the exact time and date the file was created to prevent any conflicts. If a file were being written to, it was decided the file name should contain a unique character at the start. This had not been decided at this point.

### Data Logging (Data\_logger.py) Design

The data logger’s intention is to retrieve readings from the GPS and air quality sensor write these to logs files. The functionality of the python script was not difficult to understand, so it was believed that not many functions would exist in the script.

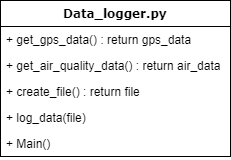


Figure - Data\_logger.py functions

The functions shown in Figure 14 were expected to provide the required functionality. The "main()” function would be used to call other functions in the correct order. This meaning the “create\_file()” would be called when creating the file before logging any data. The two functions that would return the required data would then be called, followed by the “log\_data()” function, this would store the data that had been collected.

## User Interface Design

## Other Relevant Sections

# Implementation

# Testing

## Overall Approach to Testing

## Automated Testing

### Unit Tests

### User Interface Testing

### Stress Testing

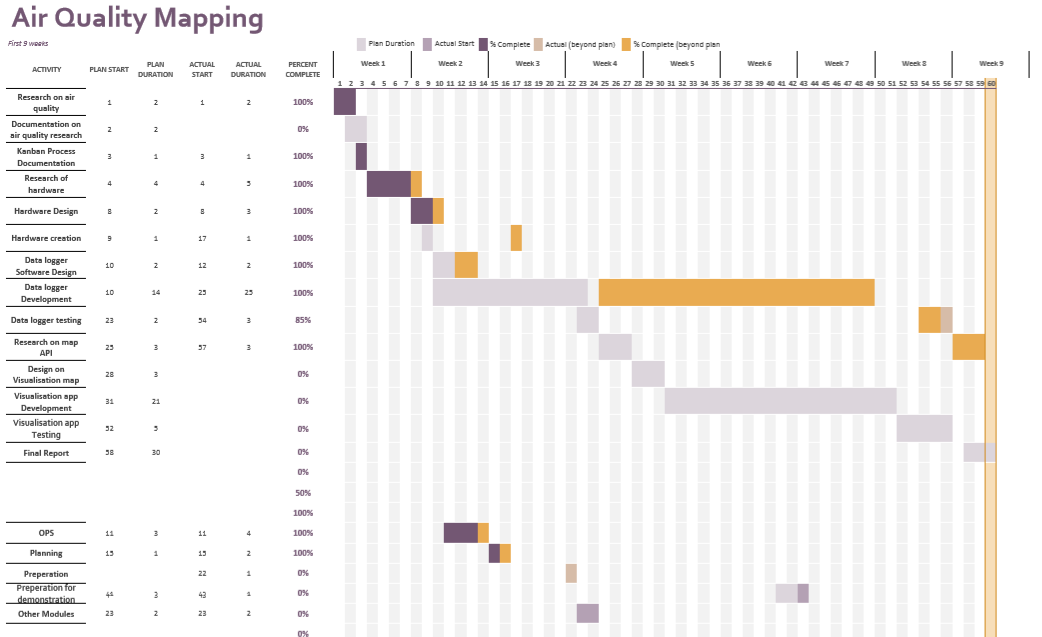
### Other Types of Testing

## Integration Testing

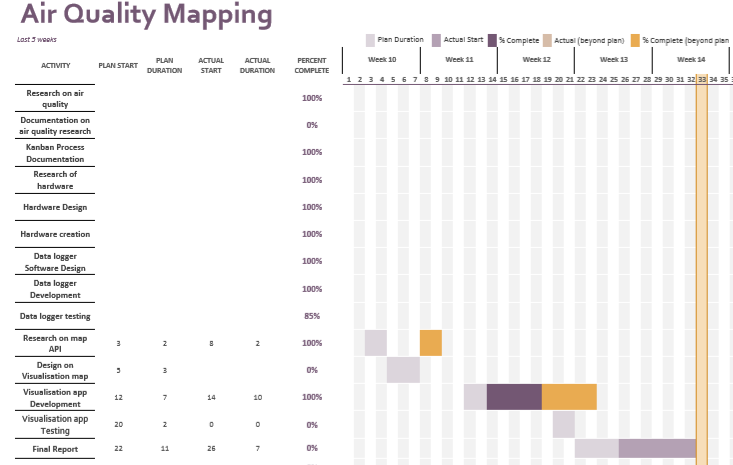
## User Testing

# Critical Evaluation

# Appendices



Appendix



Appendix



Appendix



Appendix

|  |
| --- |
| Timestamp,Latitude,Longitude,pos\_fix,CO2,TOC  1234,56.52,254.2,1,255,655  1235,56.52,254.2,1,256,664  1236,56.52,254.2,1,255,655  1237,56.52,254.2,1,256,664 |

Appendix

* 1. Third-Party Code and Libraries
  2. Ethics Submission
  3. Code Samples

# Annotated Bibliography