**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  
http://www.pigeonairpatrol.com/

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.

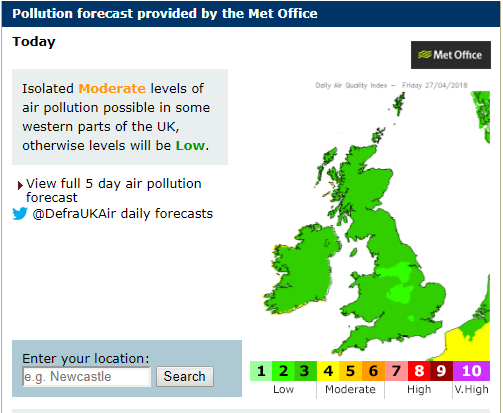


Figure - UK-air DEFRA screenshot  
https://uk-air.defra.gov.uk/

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.

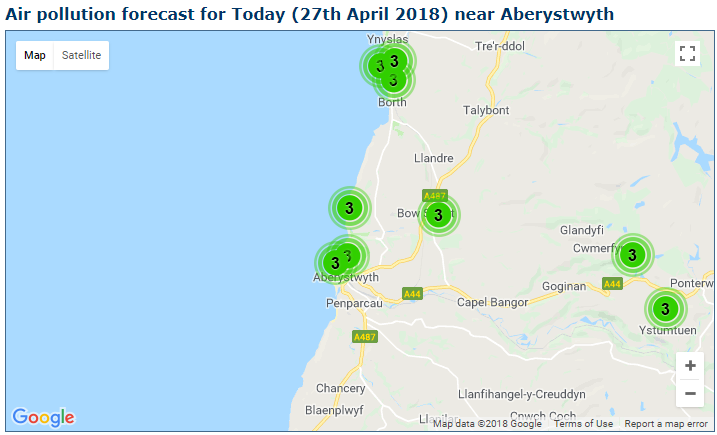


Figure - interactive map on UK-air DEFRA  
https://uk-air.defra.gov.uk/

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.

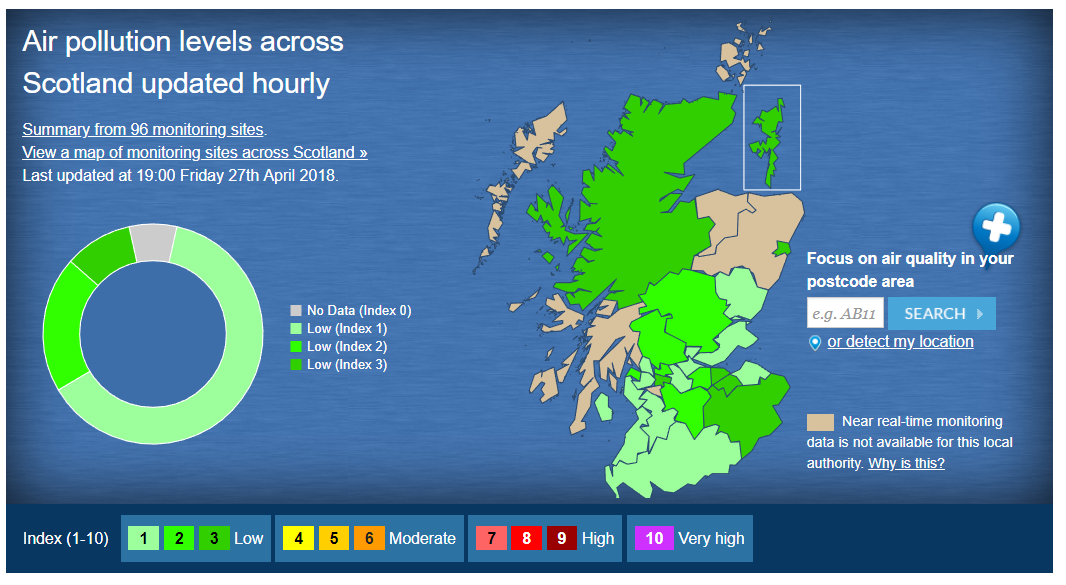


Figure - Interactive welcome screen Scottish Air Quality  
http://www.scottishairquality.co.uk/

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.

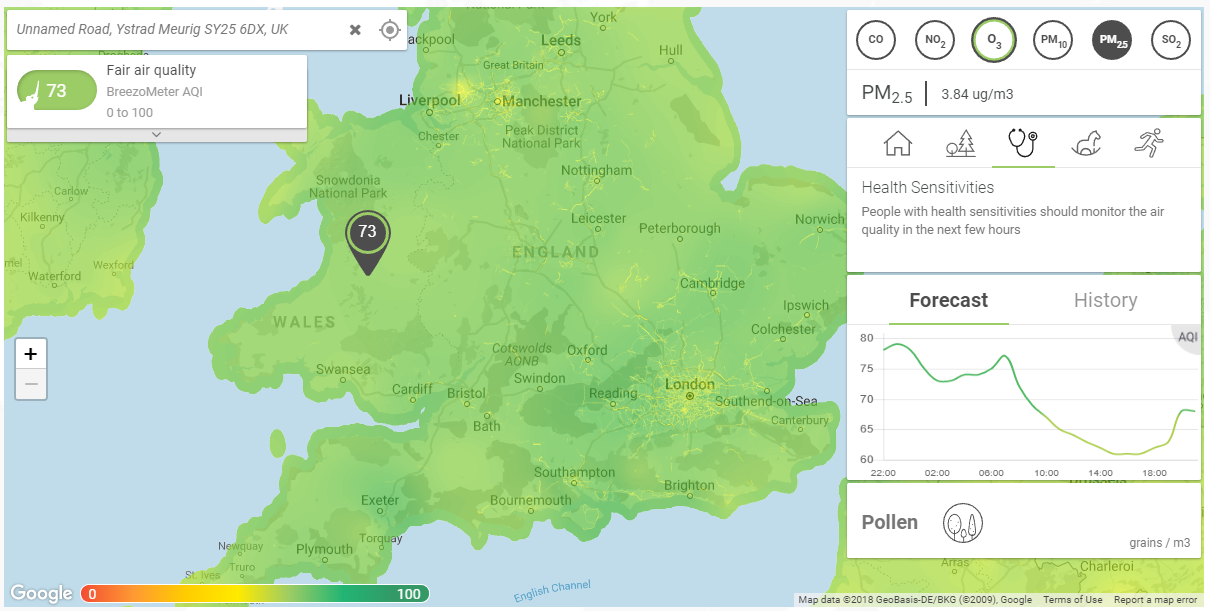


Figure - Breezometer interactive map  
https://breezometer.com/air-quality-map/

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, it shows 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 manage work without having multiple versions scattered across multiple devices. In general, it makes the process of completing the project a lot easier.

Once a task had been complete the idea was to write an in-depth explanation on an online blog to keep track of what had been completed. This was for the benefit of writing this report and for the supervisor to keep up to date.

# 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

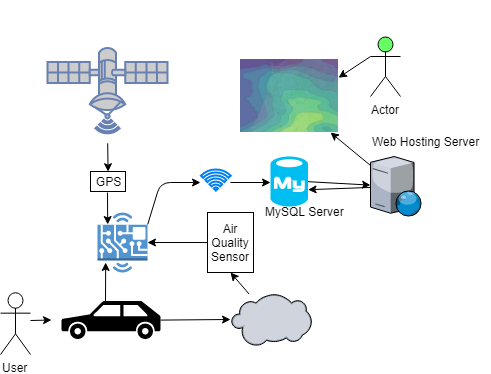


Figure – overall architecture of Air Quality Mapping

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.

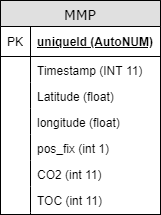


Figure - Table UML

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.

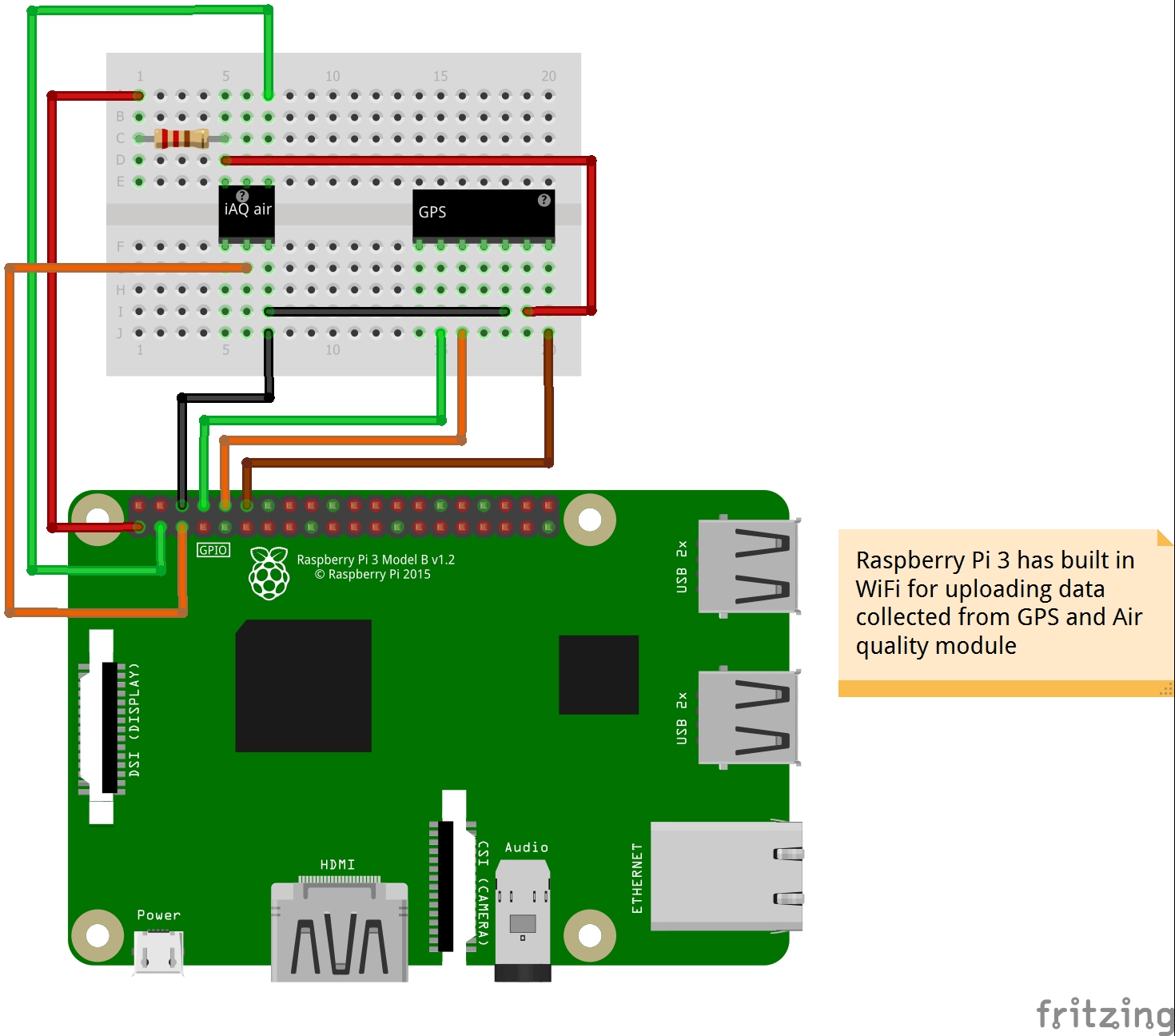


Figure - First design of Hardware

|  |  |  |
| --- | --- | --- |
| 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.

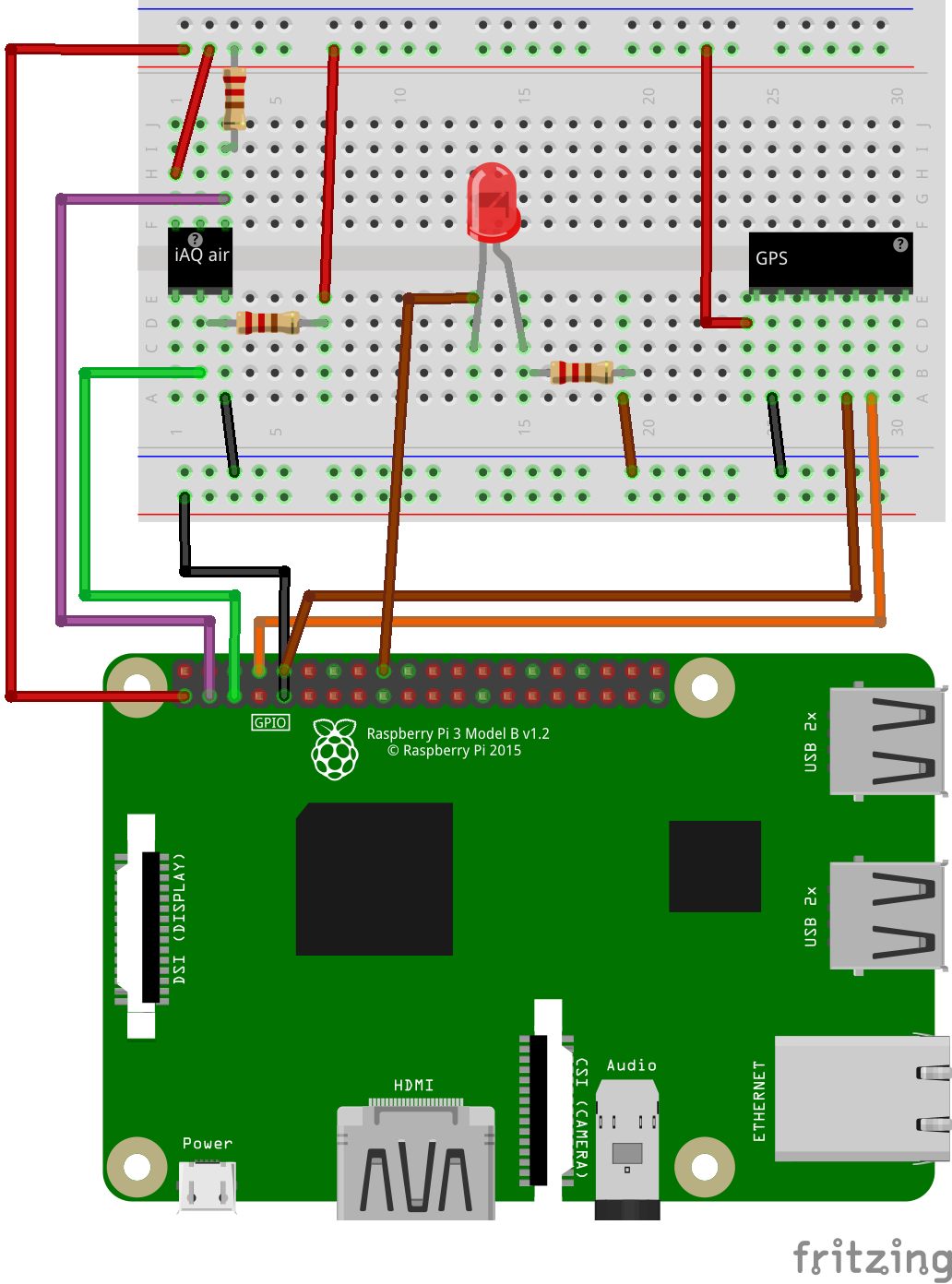


Figure - Final design for the Hardware

|  |  |  |
| --- | --- | --- |
| 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.

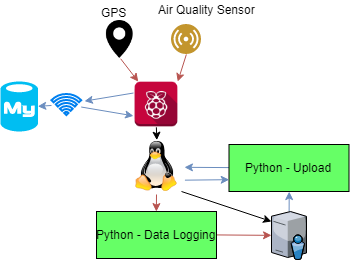


Figure - RPI Overall Software Design

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.

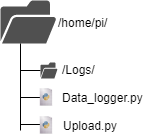


Figure - pi user file structure

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, then 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.

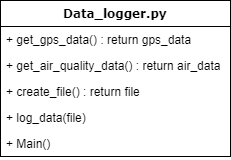


Figure - Data\_logger.py functions

The functions shown in Figure 14 were expected to provide the required functionality for the data logger script.

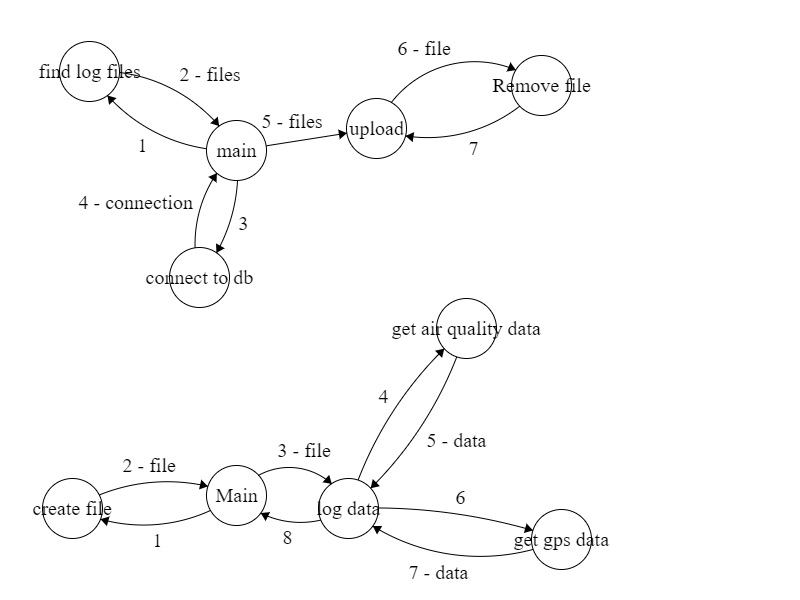


Figure - state machine showing designed function calls for data logger

The "main()” function would be used to call other functions in the correct order and would act as a loop when creating new files to log.

The first file that would be called would be the “create\_file()” function which would return a file-path or file object type. This would return back to the “main()” function. The file would then be passed to the “log\_data()” function, data would be passed back to the log data function from “get\_air\_quality\_data()” and “get\_gps\_data()” in a loop to gather multiple tuples of data to store in the file. Once enough data has been appended to the file the cycle would begin again starting at main. This process is shown in Figure 15.

### Upload (Upload.py) Design

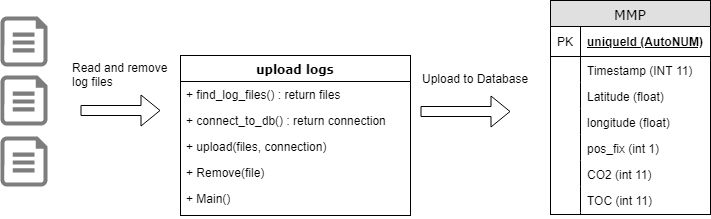


Figure – Upload.py functionality

The script will use the same programming style as the data logger which is procedurally. The scripts intention is to use the data collected by the data logging script and upload the data to the Universities MySQL server as shown in Figure 16. The design for this wasn’t really thought about until implementation but a broad design was considered.

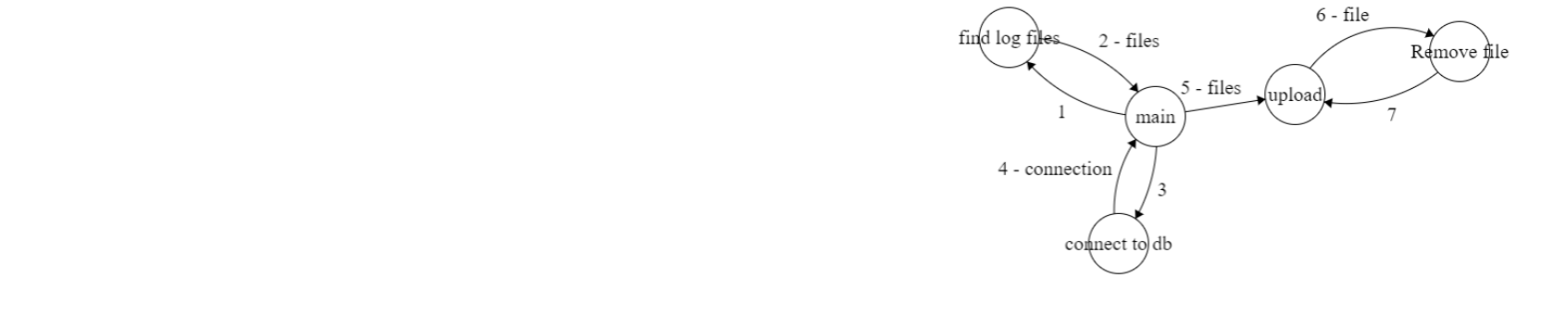


Figure - state machine showing designed function calls for upload.py

The “main()” function will provide a similar functionality to the “main()” function in the data logger script. It’s to ensure that functions are called in the correct order and sets any variables needed. The first function to be called would be the “find\_log\_files()” which would provide a list of the available files to upload in the “/log/” directory. The “connect\_to\_db()” would be called next and that would provide a secure connection to the MySQL server. The “upload()” function would use this connection to read the files and upload the data. Once the file had been completed it would be removed from the directory and the function would move onto the next file.

## Visualisation Software Design

The aim for the visualisation software was to have an online application to educate the public about air pollution on public roads. This meant that the web application had to be used to communicate with the MySQL server. To communicate with the server credentials would be necessary, these would need to be hidden from the public. Therefore, a hosting server with php was necessary.

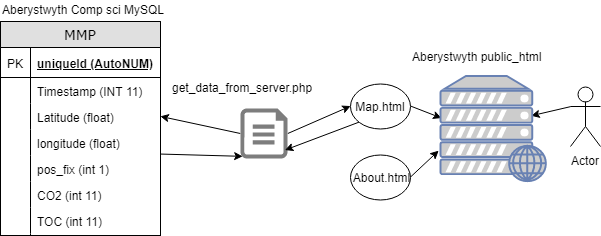


Figure - visualisation software design

Figure 18 shows the Two pages were designed to be visible to the user and these were the main page (“map.html”) and the about page (“About.html”). The user would access the pages through the Aberystwyth university student hosting server. Any php scripts could be run server side, this was designed for security in mind. The php would contact the MySQL server and retrieve all the data required for the map page.

The about page would not require any special functions (JavaScript or php). The map page however would require additional functions other than the php script. This was to handle the third party OpenStreetMap and additional plugins to draw the heatmaps. The functions would be necessary if any manipulation of data was needed for the plugin. As the web page was designed to have interactive tools to change the data, functions would be needed in this case.

## Visualisation User Interface Design

Each page on the website will contain a menu at the top of the page used for navigation purposes. The navigation options will be the “map” and the “about” pages. This will be used using an unordered list and links.

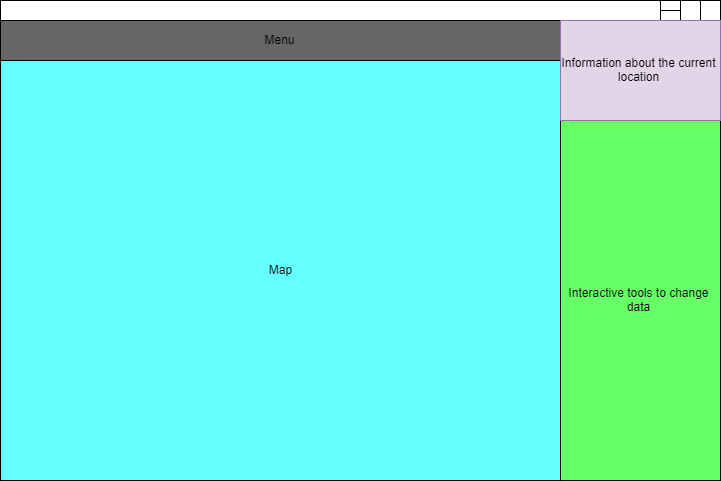


Figure - map.html page user interface design

The page was designed to have designated sections (as seen in Figure 19) this would be completed using the “div” tags provided in HTML and using CSS to allocate sections of the page. This is designed to be responsive on window size change using the “%” in CSS rather than “px”. Percentage will take a percentage of the page rather than use a pre-allocated number of pixels.

The map page was designed to be the home of the website. A large proportion of the page was dedicated to the embedded OpenStreetMap world map. This would allow users to interact and view anywhere in the world. There would be an overlay on the map controlled by a plugin that would provide the heatmap.

Pollution level Information about the current location of the maps would appear in the top right of the page. This was to include information on the effects to human health and the environment. It would state what levels are low and which are dangerous.

The interactive tools section of the webpage would allow the user to switch data from TVOC (TOC) and CO2 readings to be shown on the map. The interactive section would allow the user to change the data, but the ways in which the user could do this was not considered. The only thought about design was to include filtering to remove certain data points.

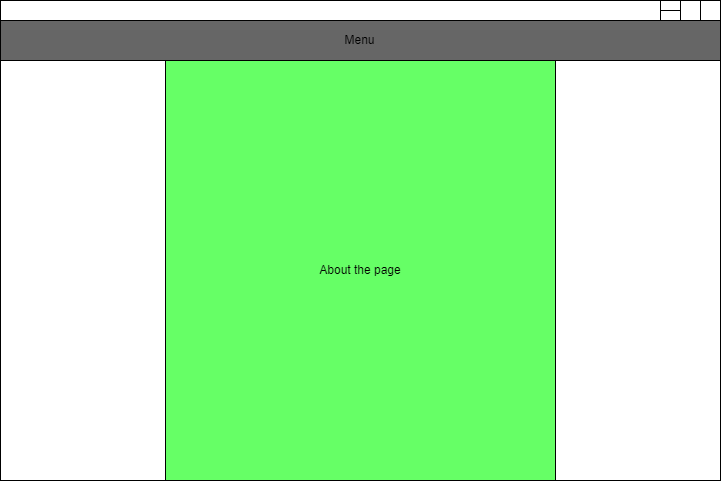


Figure - About.html user interface design

The about page was designed to display information about the website. This could include any thanks needed for libraries or installation instructions for the website. The “div” tags will be different for this page compared to the map webpage. The menu bar was designed to be 100% of the window width, stretching across the page. The CSS should centre the information on the page.

# Implementation

## Hardware implementation

When the hardware implementation had been designed and the interfaces were confirmed to work with one another, the parts were ordered.

Between the first hardware design and implementation a mistake was noticed when selecting the air quality sensor. The type of mounting had not been considered, the air quality sensor was a type of surface mount and had no pins to connect to. This was overseen as the functionality of the device was focused on. To overcome this problem a solution was found by Dr Neal Snooke, who suggested that because of the spacings between the mounts was like the spacing found on a breadboard, 2.5mm pin header strips were soldered to the air quality sensor mount spaces.

The design for the I2C interface changed when it was realised that pull-up resisters were needed (as mentioned in Section 2.2 Hardware Design). This was noticed during development of the communication between the air quality sensor and the RPI (as mentioned in Section 3.3.1 Data logger).

During development it was difficult to know whether the RPI was logging any data, this then led to the LED being implemented for the hardware. The script would cause the LED to flash a number of times depending on the current state.

## Setting up the environment for the Raspberry Pi

When setting up the environment I had decided on using the Raspbian Lite distribution of Linux. The OS image was downloaded from the Raspberry Pi website. The website mentioned the use of a SD flasher for mounting the image called Etcher. Once the installation had completed, it was connected to the RPI. The RPI was connected to the local network using an ethernet cable, the default user that is created is “pi” with the hostname of the RPI being “raspberrypi”. Putty was used in attempt to connect to the RPI to set up the WLAN using “pi@raspberrypi” as the hostname, though this failed. After researching it was found that an additional file needed to be created on the RPI in order to use SSH. The file that needed to be created on the boot partition was “ssh” with no extensions. The process was repeated and the RPI could successfully be connected to the WLAN, not needing the ethernet connection anymore. The hostname was then changed, it is advised to change the default hostname to prevent confusion if another pi was introduced to the same network, it was changed to “rdm10pi”.

As logging into the pi and editing files would no be an efficient or graphically friendly way of creating the python scripts, it was decided to set the RPI as a server to simulate the standard client-server arrangement. This took longer than expected. I installed samba through the command line using “sudo apt-get install samba samba-common-bin” which installed with no problem. The samba configuration file needed to be correct for it to work with the Windows computer on the network. The samba configuration was located in “/etc/samba/smb.conf”. A configuration (Appendix F) was appended to the end of the file but it didn’t seem to appear on the network. This lead to reading the documentation for samba. It was discovered an additional setting needed to be included to the configuration file which was “wins support = yes”, this enabled Windows support. The file share was easily mountable to the windows file explorer, this led to easily creating new files and folders within the pi users home directory. An IDE or source code editor could then be used to develop.

## Monitoring System

### Data logger

The data logger started by trying to get readings from both the UART interface on the GPS and the I2C interface on the air quality sensor. Before Serial or I2C could be implemented, it needed to be enabled on the RPI. This was done through the RPI configuration screen using the command “sudo raspi-config”.

This started by using the “serial” python library, this had to be installed using “sudo apt-get install python-serial”. It was very simple to set up. Firstly, a serial port needed to be opened, this involved using the GPS’s data sheet to find the correct settings of the serial port. This included the baud rate, whether the serial used parity or stop bits and the byte size.

Once this object was created, named “serial”, all that was needed to read a NMEA sentence on the serial port was “serial.readline()” though this was in the format of a serial message object, part of the serial library, and was difficult to access the important parts of the message, such as longitude and latitude. To transfer the message into a string, the message must be decoded into a string encoding such as “UTF-8” using the inplace function “.decode(‘utf-8’)”. To easily access parts of the serial message the message was split up into a dictionary (a dictionary is an associative array, they are indexed using keys of any data type) using a string as the key. The string chosen for each value represented what the value stood for. The dictionary was created using two lists, a list of keys, being labelled strings, and serial sentence. The serial sentence was one string containing all the values, but using “.split(‘,’)” would split the string into a list every occurrence of a comma. For an example of this see Appendix G.

It was decided that when getting data from the GPS, a specific sentence would be required before returning the dictionary. This sentence was a “GPGGA”, this was chosen due to the contents of the data. The most important values to retrieve from the GPS were longitude and latitude for this project. GPGGA contained these values along with a position fix indicator that was used to identify whether the GPS had a fix or not.

Late in to the project a mistake was noticed when converting the latitude and longitude values. It was presumed that the values needed to be divided by 100 to get the correct format of decimal degrees. The format of the data was given in degrees minutes seconds (DDMM.MMMM) which is a string of two numbers concatenated. The degrees needed to be separated from the minutes seconds and a calculation needed to be performed.

Before implementing the I2C interface, logging was introduced. Writing to a file in python is very easy with required functions being in the python core package. Information from the GPS was stored in a file, but when viewing the file blank lines would appear between each line of information. This was due to the decoding of the serial sentence to a string and the function used to write to the file. These two functions would add ‘\r\n’ causing a blank line between each message. Headings were added to the file on creation. Creating the file and logging were both similar to the design in the sense of functionality but had different variables than expected or were called in a different order.

At this stage I realised that the implementation was already deviating from the design with required functions. Two functions had been created “get\_and\_translate\_gpgga(serial)” and “set\_up\_serial()” instead of the one function in the design. This wasn’t a problem though due to the process being very focused on software development rather than documentation and requirements. Kanban allowed for easy change in requirements and design with new cards being easily added to the board.

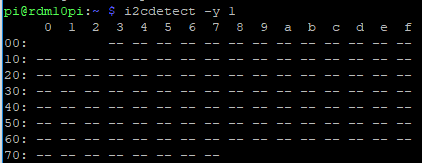


Figure - Detecting I2C interfaces

Creating communication for the I2C device was difficult compared to the serial from the start to finish. Firstly I used the Linux command “sudo i2cdetect -y 0” but the I2C default bus on the RPI 3 changed to bus 1 so this caused an error. Once I changed the bus I was given a table showing all available addresses. This worked by sending a signal to the address and checking for a reply (Figure 21).  
  
The datasheet stated that the device needed a very stable 3.3v otherwise there would be no output. As the I2C was connected to the 3.3v from the RPI it wasn’t thought that this would be the problem, but experimentation was necessary. A voltage divider was created from the 5v output of the RPI to have an output of 3.3v. This didn’t change anything, so the circuit was reverted back to its original state.

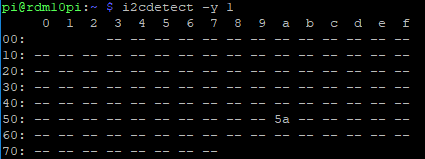


Figure - Successfully detecting I2C interfaces

After many attempts it appeared to be a connection issue, not that wires were in the wrong place, but that pull-up resistors were needed. The company that manufactures the air quality sensors also produce a click board with the same air quality sensor. A datasheet was shown with the click board and an electrical schematic was available [XXX]. The schematic showed an implementation of the I2C interface, including pull-up resistors. This showed the values required for the pull-up resistor which was 4.7K. Resistors with similar values were used and an output was successful (Figure 22).

Now that the RPI was correctly communicating with the I2C interface, software development could begin on receiving measurements from the sensor. SMBus (System Management Bus) functions were used to attempt to request data from the air quality sensor. The SMBus functions were part of the SMBus library. A small python script was created to test interaction with the I2C device. Once the bus had been configured using simple python command, requesting data was attempted (Appendix H).

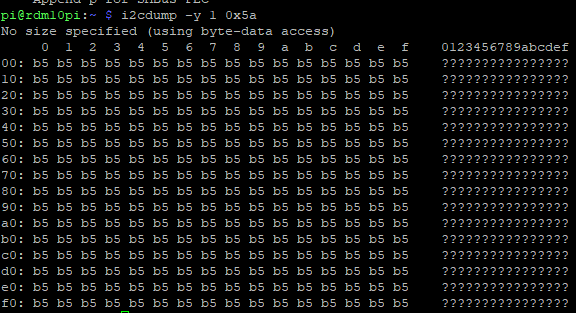


Figure - i2cdump on Raspberry Pi

When requesting the data from the air quality sensor the only response, not matter which byte was requested would be “181” or 0xB5 in hexadecimal. After looking at the datasheet for the air quality sensor it was noticed that the start of each message, the first byte would be 0xB5. When asking for multiple bytes each byte would still only contain the value 0xB5. “i2cdump” is much like the i2cdetect function, it displays all available registers through the I2C bus. Once “i2cdump -y 1 0x5a” was executed on the command line it printed all available registers (Figure 23). An attempt was made to read all the bytes available on the I2C bus address, but this made the RPI completely unresponsive.

Different functions within the SMBus library were used to attempt to request data from the air quality sensor but none worked. Research began on possible solutions. On the stackexchange forum [XXX] a post was found of a person with the same component having the same problem. The problem was caused by the I2C device resetting it’s internal state when a start condition is seen. The solution to the problem was using a different module called “pigpio”.

D:\uniwork\Git\Major-Project\Documentation\02 - Design\Software Design\Capture.png

Figure - Unusual output of byte array

Once the SMBus library was replaced by the pigpio in the interface testing script, reasonable results were being returned from the device and printed to the console. When 9 bytes were requested and where printed to the console in a byte array unusual symbols would appear that weren’t hexadecimal. “x1ax” and “0x00}” were not hexadecimal numbers so it made it difficult to use these values, especially when these contained figures needed for the air quality. It was believed that this was due to printing the results before closing the I2C connection. This however was not the case.

D:\uniwork\Git\Major-Project\Documentation\02 - Design\Software Design\Capture-1.png

Figure - pigs I2C success

As pigpio was a library for the RPI and as well as a module for python, this allowed commands to be run through the terminal. Using “pigs” commands I was able to get a reasonable result without fail (Figure 25). The problem seemed to be occurring in the python script.

After consulting with users on the raspberry pi stackexchange [XXX] it appeared to be how python would display a byte array. If the value of a byte is the same as a printable character, it would display the printable character instead. The “x1ax” would actually be two separate elements, such as “\x1a\x78\”. Once that was cleared up the byte array was accessed in the same way a normal array would be.

The software to communicate with the I2C bus could then be implemented along side the serial in the data logger. The air quality sensor has a warm up state, this is shown in one of the bytes depending on the value. When requesting information from the sensor if it is in warm up mode then another message will be requested after a short period until the warm up state has finished and the sensor is returning values.

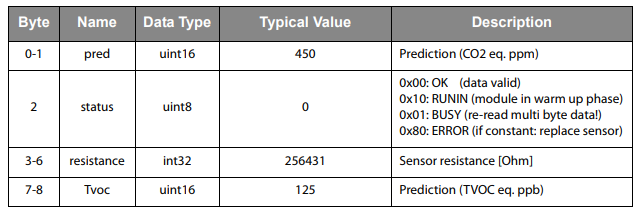


Figure - Format of bytes (Air Quality Sensor)

While implementing it was noticed that the CO2 and TOC bytes were in big endian format, a function was created to extract the data from the array and create a value from the two bytes before logging the values. The bytes in position 3-6 were ignored as this is just the resistance across the sensor, after checking the datasheet, this would not affect the TVOC or CO2 readings.

When readings were able to be logged from both sensors, the file that the data was being logged to would continuously grow. This was thought to be a bad idea as a lot of data relies on one file not corrupting. Data was then stored in maximum sized files of 25kb. This led to another problem of not knowing which file is currently being written to. To resolve this issue, while a file was being used the filename would start with “~”, like temporary documents. Another function was then added to prevent the build up of files starting with “~” if the RPI was turned off during logging. The function “check\_previous\_files()” is called before logging any results. The function searches through the logs directory for files beginning with “~” and will remove the character.

The first prototype had been created for the data logger. The RPI needed to start logging when it was switched on as no GUI would be provided when collecting data. This caused frustration as it was difficult to know whether the device was able to log any data or was waiting on the GPS and air quality for data. This is when the LED was implemented in the hardware. An “led\_out(flashes)” function was created to control the GPIO that the LED was connected to. At different stages in the script the LED would output several flashes to identify the user what the current state is. To ensure that data would be collected on start-up of the RPI a cron job was created. Cron jobs are used for scheduling tasks at required times, this is used to schedule jobs in most unix systems, in this case it would be to start the logger on reboot. “crontab -e” on the command line would allow a cron job to be created. To run the logger “@reboot sleep 20 && /home/pi/startlogger.sh” was appended to the end of the file.

Overall the software was not complicated to get readings from either the serial or the I2C bus. It took a lot of preparation with the hardware and time to use the libraries in the correct way. This caused the data logger for the monitoring system to take a lot longer than expected.

### Uploading data

The python script to upload data (python.py) needed to check for an internet connection and then upload the data collected by the logger to the Aberystwyth Universities MySQL server.

The first function created was to check for an internet connection. It’s easier to check for a network connection rather than an internet connection so this had some thought about it. The simplest method of doing this was to ping an online target. This was done using a socket library, this would allow access to the socket interface and would create a connection to an address. While implementing the script the address “www.google.com” was used. The function would return true if a connection was valid but false if there were an error.

At this point the script needed to connect to the MySQL server. It was soon realised that to access the server a local on-site connection needs to be made, or a virtual private network (VPN) be set up. As the University had strict rules on what could connect to the network this left one option. The VPN had to be set up on the RPI.

Firstly, an attempt was made to use a Linux package called “pptp” meaning point-to-point tunnelling protocol. Through the terminal to test the VPN connection the following command was executed:

*sudo pptpsetup –create abervpn –server vpn.aber.ac.uk –username rdm10@aber.ac.uk –password \*\*\*\*\*\*\*\*\*\*\*\* –start*

but this issues an authentication error. It was thought that this was due to the wrong protocol being used, that the university didn’t use a point-to-point tunnel protocol. Communications with information services at the university suggested that OpenVPN be used as this works for unix based systems. OpenVPN is supported on the RPI so pptp was removed from the RPI and OpenVPN was installed. OpenVPN requires a configuration file with certificate keys and various other parameters to start the VPN. The windows client that information services provides for connection to the VPN had the configuration zipped inside of the .exe. Once this was extracted and moved to the correct directory “sudo openvpn Aberystwyth.opvn” would appear to connect successfully to the VPN. To run this as a background process the “—daemon” tag was needed. Though the VPN stated it was successful, a test was conducted to ensure it was working. This involved using the command “curl http://ipecho.net/plain” to return an ip address before and after the connection had been made.

To run the VPN in python a module was needed called subprocess, this allowed unix commands to be run in script. It was very simple to use, the commands needed to be in an array with each command being its own element.

Now that the VPN was running and connected to the university, network connection to the MySQL server could be established in the python script. To connect to the server an additional package was used called “pymysql”, a free open source MySQL client library. The GitHub readme showed many examples of the package being used, this was very easy to follow. A connection, representation of a socket, needed to be made using the servers host name and user credentials. Instead of these details being written in the code, these were placed in a separate text file and read in, though they were still unencrypted. A cursor object would then be used to execute commands on the server. A test was connected to print all the available tables on the server, this was then implemented into a function to ensure that the table needed for the data upload existed on the database. If the table didn’t exist, then it would create the table using the cursor object. This was very easy to set up as many examples guided the process, though one mistake was made. When testing the commands, a “DROP table” command was being issued but was replaced with a command to drop all the tables from the database. This was a problem as a blog with in depth detail of processes during this project was unable to connect to the table containing all the posts. After being resolved with computer science support, back ups of the database were made on a frequent basis.

At this point a successful connection to the database could be made and if the table that was needed didn’t exist, it would be created. One last function was created and that was to construct a MySQL statement for each file in the logs directory. The files would be read one by one and uploaded to the database. The statement was built using a recursive function to add the values to the end of a string. This statement would then be executed to the server but would not be committed to the table. This wasn’t until the “commit()” command was run. This was to ensure all data from the file was uploaded to the server in case the RPI was powered off and ensure a valid transaction. Once a valid transaction has been complete then the file would be deleted. All files starting with “~” would be ignored as these are being using by the logger.

## Visualisation

# Testing

## Overall Approach to Testing

## Automated Testing

### Unit Tests

### User Interface Testing

### Stress Testing

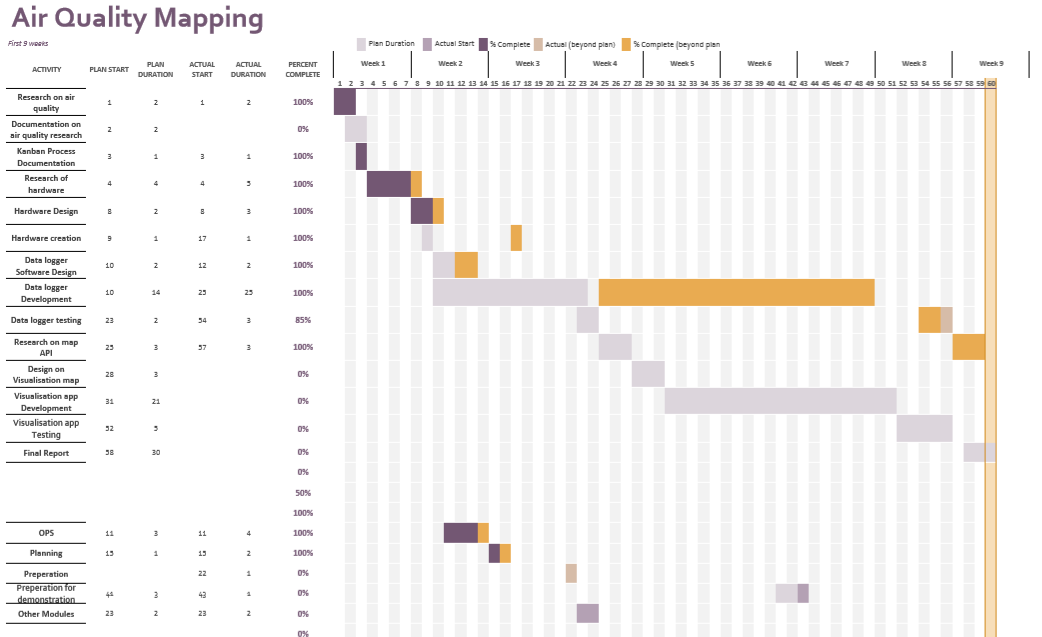
### Other Types of Testing

## Integration Testing

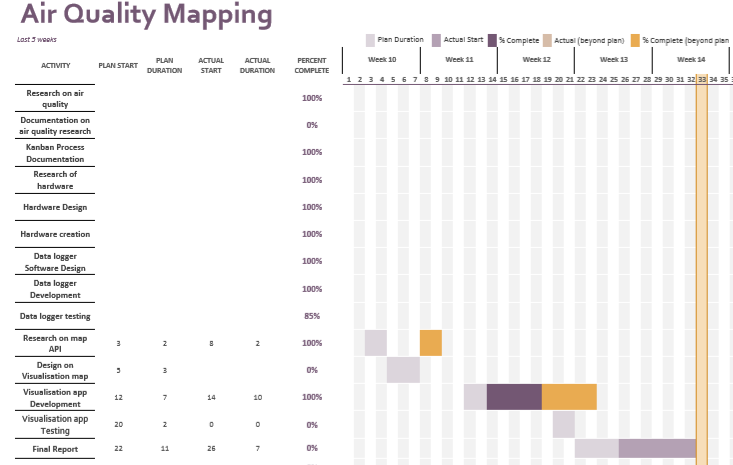
## User Testing

# Critical Evaluation

# Appendices



Appendix – first 9 weeks Gantt Chart



Appendix – last 5 weeks Gantt Chart



Appendix – Chosen air quality sensor



Appendix - chosen GPS

|  |
| --- |
| 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 – example of log file design

|  |
| --- |
| [pishare]  comment = pi dir  path = /home/pi  browseable = yes  writeable = yes  only guest = no  create mask = 0777  directory mask = 0777  public = no |

Appendix - Samba file example

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

|  |
| --- |
| message = serial.readline().decode('utf-8').replace('\r\n','')  gpgga\_message = dict(zip(["message\_ID",  "timestamp",  "latitude",  "ns\_indicator",  "longitude",  "ew\_indicator",  "position\_fix\_indicator",  "satellites\_used",  "HDOP",  "msl\_altitude",  "units",  "geoal\_seperation",  "units",  "age\_of\_diff\_corr",  "checksum"],message.split(","))) |

Appendix - reading from serial

|  |
| --- |
| import smbus  BUS=smbus.SMBus(1)  ADDRESS = 0x5a  print(BUS.read\_byte\_data(ADDRESS,0)) |

Appendix - Requesting data using SMBus

# Annotated Bibliography