Development and Deployment of Low-Cost Environmental Monitoring Sensors: Some Lessons Learned

1st Author Name, 2nd Author Name, …, nth Author Name
[LEAVE BLANK FOR INITIAL SUBMISSION]

Affiliation(s) [LEAVE BLANK FOR INITIAL SUBMISSION]
Location, Country[LEAVE BLANK FOR INITIAL SUBMISSION]
Contact Emails[LEAVE BLANK FOR INITIAL SUBMISSION]

Abstract

The aim of this paper is to describe the development and deployment of a cluster of 20 very low-cost environmental monitoring sensors, constructed from cheap and readily available, off-the-shelf components. The design of these sensors was made particularly challenging since the domain in which we were working was pervaded by numerous technical, physical, political, social and financial challenges. All of these factors result in the need to make numerous compromises and complex trade-offs over the course of the development process. We explore all phases of the project, from design and conceptualisation, all the way through to deployment and data retrieval. We discuss the challenges faced and solutions found at each step of the way and conclude with a set of "Lessons Learned" that may help to guide others attempting similar activities.

Keywords

Greenspaces, Environmental, Monitoring, Community-Based Monitoring

Purpose of Project

In this current era of limited funding for public green spaces, it is essential that informed and insightful spending takes place in order to achieve the most optimal investment of public money. Only by understanding levels and patterns of green space utilisation can we make informed decisions to maximise the benefit of such spending.

To justify our interest in the provision and upkeep of urban parks, we need only consider the public good derived from such spaces. This not only includes the obvious physical health, mental wellbeing and recreational benefits, but also the environmental benefits, including the regulation of air quality and climate, water quality protection, rainfall retention and storm management and reduction the urban "heat island" effect [4].

This paper describes and documents a project to embed sensors within urban green spaces in and around the city of Plymouth in the UK. The objective of this project was to determine the usage patterns of these green spaces in order to:

1. Provide a justification for capital investment as part of ongoing year-on-year development as well as additional one-off intervention expenditure
2. Assist in the identification of spatial hotspots to aid in the geographical targeting of resources (such as placement of interpretation signage)
3. Identification of temporal peak-times to aid in delivery of services (such as urban ranger time-on-site)
4. To provide quantitative data to support evaluation and demonstration of impact of investment and provision of resources and services
5. To aid in spatial and temporal targeting of qualitative evaluation tools (such as user questionnaires and surveys) to support evaluation and demonstration of impact

Data collection through manual observation and counting has been undertaken in the past within these spaces, however this process has its limitations. Due to the need to involve human operatives, it is time-consuming, expensive and only practical for relatively short durations of study. There was a clear need for an automated sensing mechanism that could monitor space usage patterns around the clock, for an extended period of time.

Community Participation

A key element of the described project is the involvement of the local community in the long-term monitoring and development of urban greenspaces within Plymouth. This strategy of inclusion aims to increase public perceptions of value and encourage greater citizen involvement in the stewardship of these public spaces.

Within this project, we wished to involve the community (through park rangers and "friends" of the green spaces) in both the deployment of devices and retrieval of data. As D'Hondt et al observe [2], this kind of participatory involvement in sensing projects can help close the feedback loop between the gathered data and people who live and work in the study environment. Not only are citizens crucial in assisting with large-scale data collection, but the data gathered can reflect back on citizens and nurture their awareness of the environment in which they live.

Within such domains, collaboration between Institutions, State and Community is become much more common. As Loka has noted "More often than not, community-based research involves the collaboration of community members (represented by grassroots activists, community-based organisations, etc.) and experts (represented by university researchers and professional scientists.)" [10].

Such collaborations have solidified into the notion of Community-Based Monitoring, where concerned citizens, government agencies, industry, academia, community groups, and local institutions collaborate to monitor, track and respond to issues of common community concern [13]. Legg and Nagy [9] note how such approaches can be effective in identifying when an ecosystem is departing from a desired state, detecting effects of perturbations and disturbances and measuring the success of management actions.

Conrad et al [1] identify a number of high-level benefits of Community-Based Monitoring including increased environmental democracy, improved scientific literacy, social capital, citizen inclusion in local issues as well benefits to government and the ecosystem being monitored. In order to achieve these potential benefits Conrad highlights a set of key challenges which must be overcome including data fragmentation, inaccuracy, lack of objectivity, access to collected information and ensuring the utility of data for decision-making and environmental management. We hope to address many of these challenges within the work we describe.

Design Constraints

Whilst attempting to develop a technical solution that will achieve all of the objectives outlined above, it is essential to keep in mind the various factors that constrain and impinge upon the possible range of design choices. Any successful solution must demonstrate the following characteristics:

1. Easily replicable: a substantial number of devices (20) are required to cover the numerous sites and locations
2. Low cost: the devices must be very cheap (only £2000 was available for all parts and labour for all 20 devices).
3. Environmentally robust: the sensor devices must be able to operate in a range of challenging weather conditions
4. Physically robust: the device must be resistant to theft and vandalism
5. Infrastructural independent: the devices must operate without need for technical infrastructure (for example mains power or data communication network)
6. Operationally independent: the devices must be able to operate without human intervention or administrative support for at least 2 weeks (i.e. the planned duration of the study)

Although the idea of sensor-based monitoring of green space environments is not a new one, there is a tendency for such projects to attempt a much higher level of complexity (and therefore cost). For example, we can consider the Hyde Park Sensing Project [6]. The focus of this work was the deployment of a network of wireless sensors for the monitoring of soil, air, water and footfall within green space environments. Although very relevant to our own intended outcomes, this project is far more ambitious and extravagant in terms of cost of materials, power consumption and required network infrastructure. As such, this kind of approach is inappropriate for the environment and context of our own work.

Developed Solution

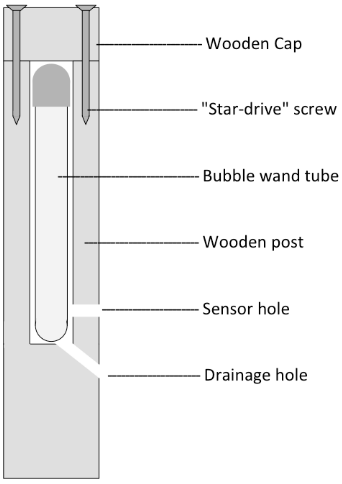
We now provide an outline description of the solution created by our development team to address the objectives and challenges described above. In order to fall within the very limited budget, it was essential to make use of cheap, mass-produced, readily available off-the-shelf components (as there was neither the time nor the money to develop bespoke equipment and components). For these reasons the core of the sensor device makes use of the "Ardulog" Arduino-based data logging board [8]. This provides a low-cost programmable microcontroller with integrated real-time clock and SD card data logging facilities. By using an SD card to log data, we remove the need for expensive and power-hungry network communication infrastructure. Although this means that the stored data must at some time be manually retrieved, this is a task that can easily be carried out by volunteers, thus helping to ensure the close involvement of the community in the monitoring process.

The actual sensor used to detect human presence and activity was a passive infrared detector (a type used in burglar alarms, automatic lighting systems and so on). These are very simple, widely available devices that sense movement by monitoring subtle changes in infrared radiation given off by warm objects (such as humans). These detectors are typically used for wide-area sensing, however we were able to achieve more directional operation by masking off the device and using a guide hole to narrow the angle of coverage. We chose not to use beam-braking or camera-based devices due to the higher power consumption of such sensors.

We made use of standard AA batteries to provide power for the whole device. Using open-source software libraries, it was possible to develop very power efficient firmware to run on the logging boards. This involved putting the device into a power saving "deep-sleep" mode, with it only being woken when movement was detected by the infrared sensors. Using this approach we would expect the AA batteries to last at least 4 weeks, in normal working conditions [7].

It proved impossible to find a single integrated, waterproof, temperature resistant, vandal and theft proof container within the limited budget. We could not identify a single case that met all of the above requirements. To solve this problem, we ended up using a composite enclosure that consisted of two separate layers (see cross-section diagram in Figure 1 below).

Figure 1. Composite Enclosure Cross-section



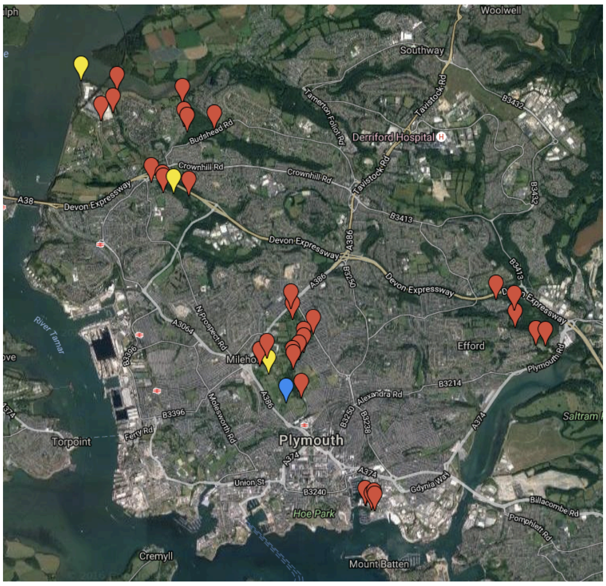
The inner layer was a thin waterproof plastic tube, repurposed from off-the-self bubble wands (Figure 2). The outer layer was a thick wooden fencing post with a deep hole drilled into it to receive the inner plastic tube (Figure 3). Additional holes were drilled in the wooden post to allow the infrared sensor to detect activity and a downward-facing drainage hole to prevent the accumulation of water inside the post. A thick wooden cap (secured by "star drive" screws) was attached to the top of the post to protect the device from the elements and from theft. The completed whole was then wire-tied to a permanent structure (e.g. existing signs or fence posts) again to help prevent theft.

Figure 2 & 3. Inner and Outer layers of enclosure

The composite two-part enclosure provided ideal protection for the sensor device. The outer wooden layer provided robust, tough and insulated protection from physical damage and temperature extremes. The inner plastic tube provided a final waterproofing layer to protect the delicate components during insertion and removal from the wooden post and to prevent humidity and condensation affecting the electronics. We did consider using a desiccant (silica gel beads) inside the plastic tube, however previous experience indicated that this could be counter productive – with the desiccant acting as a water pump (drawing moisture into the tube) if the lid is not perfectly airtight. The selection of bubble wand used in the design was crucial as its diameter was limited by maximum size of standard off-the-shelf drill bit that could be purchased.

In order to document the creation of the sensor devices and enclosure and to enable other groups to make use of our findings, we have produced an online tutorial describing the process. This has been published on the "Instructables" website (a site frequented by crafters, makers and electronics hackers) in order to promote the device to the grass-roots environmental monitoring community [3].

Figure 4. Map of Greenspaces within Plymouth



Deployment

For the deployment of the sensor devices, we chose various different green spaces in the Plymouth area (see map in Figure 4). Within each green space we selected a number of different locations to site the sensors. The selection of these locations was based on a number of different criteria:

- Entry and exit points into the green space areas
- Particular pinch points (such as gates and bridges)
- Areas of suspected congregation of green space users

An additional limiting factor was the availability of existing permanent structures (e.g. signs and fence posts etc.) against which the sensor devices could be securely fixed. A final deployment constraint was the need to set the height and the angle of the sensors in order to avoid false positive readings (due to proximity to areas such as car parks and known badger habitats).

The sensors were deployed at the indicated locations and left to collect data for a period of two weeks. After this period of time, the sensors were collected by members of the team and brought back to the lab for data analysis. Training sessions and written instructions were provided to those collecting the data to ensure that collection procedures were followed to minimise the risk of data loss or sensor damage.

Data Retrieval and Processing

In order to retrieve the data that had been collected by the sensor devices, the SD memory cards were ejected, inserted into a laptop and all the data files copied across. The raw data consisted of individual timestamps that indicate the exact time when activity had trigged the infrared sensor. The combined raw data for all 20 sensors consisted of 170k individual timestamps for the two-week deployment.

In order to convert this data into a suitable format for visualisation and interpretation, it must first be aggregated together. The raw timestamps hold little information on their own – they must be combined together into rates (e.g. triggers per minute or per hour) in order to yield useful information. This aggregation is achieved using a "bucketing" algorithm, whereby a day is split into a fixed number of timeslots (e.g. one for each hour of the day) and the various sensed timestamps are distributed appropriately in order to fill those timeslot "buckets".

Once this aggregation task is complete, the data can then be visually rendered using one of a number of different visualisation techniques. The visualisations rendered as part of the project included the following:

- Simple line graph: Height of the line above the x-axis indicates activity rates
- Linear temporal heatmap: Colour "temperature" of a bar indicates activity rates
- Radial temporal heatmap: Colour "temperature" of a 24 hour ring indicates activity rates
- Cloud clustering diagram: Size and opacity of a cloud formation indicates activity rates

Examples of each of these types of visualisation are shown in figures 5 to 8 below.

Each visualisation technique has its own advantages and disadvantages and reveals different aspects of the data. Some are useful for gaining an overall appreciation of the patterns of activity with in the green spaces, others provide more detailed, low-level insight into visitor numbers.

Figure 5. Simple Line Graph

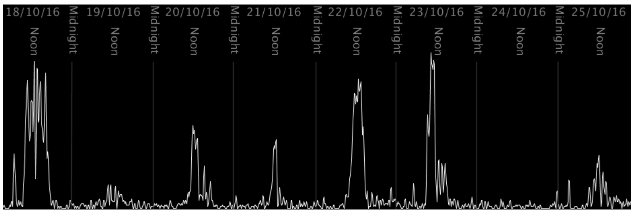


Figure 6. Linear Temporal Heatmap

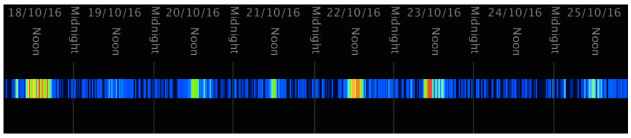


Figure 7. Radial Temporal Heatmaps

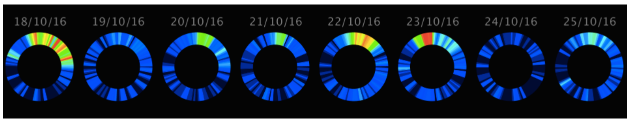
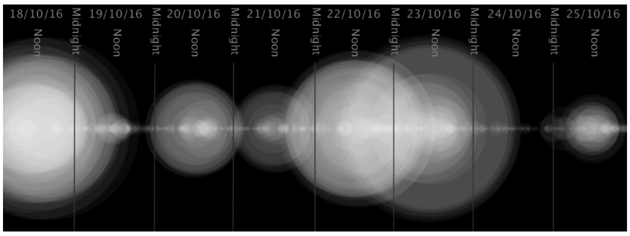


Figure 8. Cloud Clustering Diagram



Observations and Findings

In this section we present a number of observations relating to the various phases of the project. In particular we discuss the successes and failures of the various design decisions made and strategies that we have employed; we consider the accuracy of the data collected; finally we examine the role of visualisation in the interpretation of the data.

Failure Rates

Of the 20 devices deployed we experienced an attrition rate (failures, damage, theft) of approximately 20-25%. This number was very much within the range we expected and included:

- One device with wiring problem that failed shortly after installation
- One device with installation problems that produced a constant stream of false positive data
- One device with a failed clock battery (the clock became reset to 1st Jan 2000)
- Two posts that had been vandalised and the sensor devices stolen/destroyed

No problems with condensation or extremes of temperature were experienced, however the temperature range during the study was not particularly wide, varying as it did between 6-20 degrees Celsius.

The construction of the composite enclosures were robust and the wooden posts proved rugged and damage resistant. The two devices that were vandalised where those that were not as securely anchored to permanent structures as the rest. One factor which is likely to have contributed to the survival of the majority of the devices was an attempt to "normalise" their appearance. It was noted during the installation process how new and clean the posts looked. This would have undoubtedly attracted the attention of passers-by to these new additions to the green spaces. By rubbing mud into the posts once they were in situ, we are able to better blend them into the environment and make them much less conspicuous.

Data Accuracy

One of the main concerns that arose during the project was the issue of data accuracy. Hondt et al observe that the main hurdle to be overcome with most participatory sensing projects is that of data quality. Homemade, low-cost sensors are frequently less accurate than professionally equipment and the key question raised is whether the volume of data collected can compensate for the inherent lack of accuracy. In the case of our project, the main source of inaccuracy was the reliability of the infrared sensors and the false-positive triggering of the devices by a range of environmental phenomenon. These included dogs being exercised, badgers at night, the glare from the sun, people loitering near the sensors, and even people playing "Pokemon go" in the parks.

It is essential to ask ourselves what impact this data inaccuracy has had on the outcome of our study. Clearly it is important that we are aware of the limitations of the data and do not attempt to interpret it purely at face value. It is important to take a pragmatic approach to working within the constraints of the study and we have learned to embrace the incompleteness and inaccuracies in the derived data sets. Provided that we are aware of the deficiencies within the dataset, we may still derive utility from the raw data. As long as we accept it as an *indication* of activity levels within the green spaces (rather than an absolute measure of footfall) we can infer useful information from it.

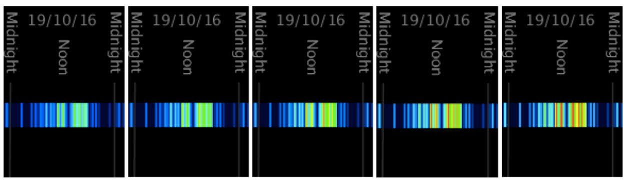
In a similar study D'Hondt describes how the systematic and scientific calibration of cheap monitoring devices can greatly improve the accuracy of collected noise pollution data [2]. In so doing, it is possible to collected data of equivalent (or even superior) quality to professional and scientific gathering techniques. It is likely that a similar approach to calibration could be developed for use with our own work. Although beyond the scope and timeframe of our project, we did undertake a short period of small-scale manual counting as an attempt to provide a rough validation of the data collected by the sensors. By extending such activities, it would be possible to derive a number of calibration formulae that could allow extrapolation of more precise footfall data from the raw motion/activity data. This however is very much future work, with much more effort required to develop these formulae for a range of locations and conditions and to assess the level of accuracy of the derived results.

Even without such calibration, we argue that valuable information may still be derived from the raw gathered data. In their paper on volunteer monitoring of water quality, Savan and Gore observe that patterns and changes are often the most useful (and achievable) results. Participatory monitoring groups often focus on biological indicators that provide a warning of problematic water quality, rather than rigorous and precise measures water chemistry. In a similar way, our focus on abstract patterns of behavior (rather than absolute measures of footfall) can provide valuable insights into the usage of green spaces. This is aided by suitable visualisation techniques that support investigation and discussion of such abstract patterns.

Data Visualisation

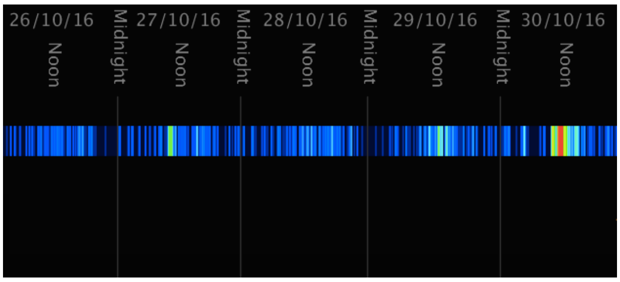
The range of previously illustrated visual representations has allowed us to abstract over the low-level inaccuracies of the raw data and permitted overall patterns to be observed. Exploration of these visual patterns (in particular through the use of heatmaps) enabled some interesting discussions to take place between the technical support team and the green space managers. It is important to note that the described visualisations are not just static images, but are interactive in nature, with adjustable degrees of aggregation and controllable levels of contrast. To illustrate this, the screenshots in Figure 9 below show the same day, but with varying levels of contrast.

Figure 9. Varying levels of Contrast

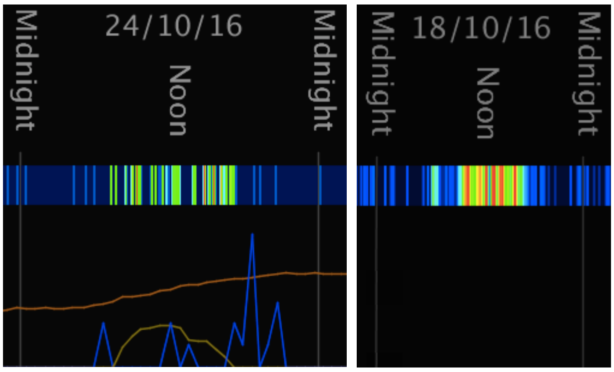


Using such interactive visualisations, we can manipulate the settings to reveal interesting patterns within the data sets. Observable patterns include expected behaviours, such as peak times on Sundays (see 30/10/16 in Figure 10) when many people go out for a walk in the park; the impact of wet weather on people desire to be outside (Figure 11 shows correlation between heatmap and the precipitation graph in blue below). In addition to these, there were also various unexpected behaviours, a prime example being the "Mysterious Monday" segment shown in Figure 12 below.

Figure 10. Popularity of Sunday



Figures 11 & 12. Impact of rain & "Mysterious Monday"



This "Mysterious Monday" segment is worthy of further discussion. It was first identified during data retrieval and investigation in collaboration with the green space managers. It illustrates an unusual anomaly at what is typically a quiet location, during a quiet time of the week. No official event had been planned and no one on the team was aware of anything unusual taking place within the park at that time. After an interesting discussion over the possible causes of this anomaly, it was concluded that it was most likely to have been a school field trip (or similar event) visiting the park that day. What the "Mysterious Monday" segment does illustrate is the rare insight into behavioral patterns of usage that the gathered data offers us. Insight that would not have been possible if it were not for the sensor devices deployed within the various green space locations.

Lessons Learned

In this concluding section, we draw upon all of the experiences gained during this project. It is hoped that the insights gleaned may help others attempting similar work. We summarise these experiences in the following concise list of 9 lessons learned:

1. Construction can often be achieved entirely from cheap, readily available off-the-shelf components. "Cheap" does necessarily equate to "badly designed", although it often requires more creativity and ingenuity to work within tight budgetary constraints.
2. The design of effective enclosures is by no means a non-trivial task should not be overlooked. Time and money should be invested in creating them and this task should not be taken for granted nor left until last.
3. Training is crucial to the success of volunteer-supported deployment, recovery, data interpretation and visualisation. Without suitable support there is the danger that inaccuracies will be introduced, data lost or delicate equipment damaged.
4. Anticipate loss and failure because it will almost certainly occur. Incorporate strategies to deal with such failures, for example ensuring redundancy and designing for modular replicability.
5. Manage expectations of decision-makers and sponsors regarding anticipated accuracy and completeness of the final data set. Ensure that they are made aware of the anticipated quality of the data as development progresses, so that there will be no disappointments at the end of the project.
6. Focus on capturing relative measures that flag change, rather than attempting absolute measures of questionable accuracy. It is better to derive simple, reliable data rather than trying (and failing) to measure the impractical.
7. Suitable aggregation, abstraction and visualisation tools are crucial to extracting useful information from dirty data. Without such features the holes within the data will obscure the informative patterns that do exist.
8. Systemic calibration is a potential route to deriving more accurate absolute data (if it is truly required). It is important to ask ourselves whether such data is truly essential in order to achieve the objectives of the study.
9. Urban sensing is often just as much a sociological challenge as it is technical or environmental one. We cannot disconnect the physical environment from its social context and must therefor design our devices and protocols to deal with issues such as community buy-in, volunteer ability, device theft and vandalism.

As a final observation, we reflect on the fact that some of the activities involved within the described project were manual and labor-intensive in nature. The various physical, technical and financial constraints placed upon the project meant that it was impossible to develop fully autonomous and independent sensor devices. The main two technical challenges that stood in the way were power consumption and data communication. It is interesting to note that some exciting developments within both of these areas are currently being explored within the wider research and development domain. This includes energy harvesting, with technology such as FreeVolt [5] and low-power, long-range communication technologies, such as LoRa [11] and Sigfox [12]. Clearly there is likely to be some interesting innovations in the area of low-cost environmental sensor devices in the near future when such technologies become cheaper and more widely available.

References

[1] Conrad, C.C., Hilchey, K.G., *A review of citizen science and community-based environmental monitoring: issues and opportunities*. Environmental Monitoring and Assessment, May 2011, Volume 176, Issue 1, pp 273–291.

[2] D'Hondt, E., Stevens, M., Jacobs, A. *Participatory noise mapping works!*. Pervasive and Mobile Computing, Volume 9, Issue 5, October 2013, pp 681–694.

[3] Digital Art and Technology (2017), *Green Space Activity Monitor* [online] Available at: http://www.instructables.com/id/Green-Space-Activity-Monitor/ [Accessed 9th January 2017].

[4] EverGreen (2017),*The Environmental Benefits of Green Spaces* [online] Available at: https://projectevergreen.org/wp-content/uploads/2013/07/EnvironmentalBenefitsofGreenSpace.pdf [Accessed 9th January 2017].

[5] FreeVolt (2017), *FreeVolt – Low Energy IoT* [online] Available at: http://www.getfreevolt.com/ [online] Available at: http [Accessed 9th January 2017].

[6] Future Cities Catapult (2017), *Hyde Park Sensing Project* [online] Available at: http://smarterlondon.co.uk/case-studies/hyde-park-sensing/ [Accessed 9th January 2017].

[7] Crijns, H. (2017), *How to run an Arduino on AA Batteries for over a year* [online] Available at: https://hwstartup.wordpress.com/2013/04/15/ [Accessed 9th January 2017].

[8] Hobbytronics (2017), *ArduLog-RTC Data Logger* [online] Available at: http://www.hobbytronics.co.uk/ardulog-rtc [Accessed 9th January 2017].

[9] Legg, C. J., & Nagy, L. (2006). *Why most conservation monitoring is, but need not be, a waste of time*. Journal of Environmental Management, 78, 194–199.

[10] Loka Institute (2017), *About the CRN: What is Community-Based Research?* [online] Available at: http://www.loka.org/crn/About%20CRN.htm [Accessed 9th January 2017].

[11] LoRa (2017), *LoRa Alliance, Wide Area Networks for IoT* [online] Available at: https://www.lora-alliance.org/ [Accessed 9th January 2017].

[12] SigFox (2017), *Sigfox - The Global Communications Service Provider for the Internet of Things (IoT)* [online] Available at: https://www.sigfox.com/ [Accessed 9th January 2017].

[13] Whitelaw, G., Vaughan, H., Craig, B., & Atkinson, D. (2003). *Establishing the Canadian Community Monitoring Network*. Environmental Monitoring and Assessment, 88, 409–418.

Acknowledgements

The authors would like to acknowledge the help and cooperation of Ashley Tod and Zoe Sydenham from Plymouth City Council as well as Rob Stenhouse and Shaune Oosthuizen from Plymouth University.