CHAPTER 1 - INTRODUCTION

1.1 Study Background

According to the World Health Organisation there is an estimated 285 million people worldwide who are visually impaired (World Health Organisation, 2014). One of the biggest obstacles facing visually impaired people wanting to live independently concerns their ability to travel by themselves (Brock, et al, 2015). Travelling is important for someone to participate in society (Paajala & Keranen, 2015) and nearly half of visually impaired people feel like they are cut off from the society around them (RNIB, 2015). Problems with wayfinding can restrict people to travelling routes that they know or can even mean they do not travel at all (Brock et al, 2015). The impact can depend on the type and severity of the visual impairment for example reading a road sign may be difficult for someone with a loss of central vision (Bradley and Dunlop, 2005) but they may be able to see the buildings they are passing. For this research the term 'visually impaired person' is going to refer to a person who has problems navigating independently due to loss of vision.

With the invention of technologies like Global Positioning System (GPS) there has been new tools created to help visually impaired people travel independently (Brusnighan et al, 1989). These tools require geographic data that is relevant to somebody with a visually impairment. Despite research into these data requirements, for example Chandler & Worsfold (2013), there has been little research into whether different data sources meet these requirements. This research is going to look at whether OpenStreetMap (OSM) data could meet these requirements by looking at the data quality of the OSM dataset.

1.2 Study area

Though it is hoped that this research will be useful in other places, the ground survey and national analysis focuses on the United Kingdom (UK). The UK has large urban areas and in 2015 there was over two million people in the UK who had sight loss (RNIB, 2015). There is legislation that protects the rights of people who have sight loss, such as the Equality Act 2010 which can be applied to them being able to access urban spaces for independent travel (Norgate, 2012).

The UK has also been home to research into the quality of Openstreetmap (OSM) data. This includes research by Haklay (2010) into data quality in London and England and Mooney et al (2010) who looked at different areas of the UK along with other regions while attempting to develop some quality metrics.

The research focuses in on three cities within the UK, namely Bristol, Cardiff and Manchester. Urban areas are where visually impaired people tend to live and travel (Gaunet and Briffault, 2005) and so this is where research is most appropriate.

Bristol and Cardiff are both cities known by the researcher and where they have guided visually impaired people. Therefore they have knowledge of locations that a visually impaired person might visit and the routes they are likely to take.

Bristol, Cardiff and Manchester are all urban areas within the UK. In Bristol there are 1425 adults who are registered as blind (Bristol City Council, 2017). All three cities have several transport hubs and Royal National Institute of Blind People (RNIB) offices.

1.3 Dissertation Structure

In chapter two a literature review looks at the current research into the geographic data requirements for a visually impaired person, different data sources that could meet these requirements, data quality and how data quality has been looked at in regard to Volunteered Geographic Information (VGI). The data and methods that are going to be used in this research are then introduced under the different data quality elements identified in the literature review. The results are then presented, again under the different data quality elements, followed by a discussion of what the results mean. The dissertation ends with conclusions from the project and recommendations for further study.

1.4 Aim and Objectives

1.4.1 Research aim

The aim of this study is to evaluate OpenStreetMap (OSM) as a geographic data source for visually impaired people. It will focus on the data needed for wayfinding and route selection.

1.4.2 Research objectives

- Evaluation of current research into geographic data requirements for visually impaired people
- Evaluation of the elements and attributes within OSM that could meet these requirements
- Identification of suitable methods to assess data quality of OSM data for these requirements
- Analysis of the data quality of elements and attributes within OSM to meet the above requirements

CHAPTER 2 - LITERATURE REVIEW

2.1 Introduction

To understand previous research into this area a literature review was undertaken. This literature review starts by looking at the problems a visually impaired person has travelling independently and the solutions that have been found to help them. It then looks at what data is needed for these solutions, the data sources that exist and explores why VGI might be useful for this purpose. It ends by looking at different elements of data quality and how they can be used to look at volunteered geographic information (VGI).

2.2 Independent travel for a visually impaired person

For someone to travel by themselves they need to be able to navigate around their environment which Montello & Sas (2006) divide into wayfinding and locomotion. Wayfinding is where a person is trying to reach a destination that is further away than their local surroundings, for this they need to know what their destination is and the method they are going to use to get there. Locomotion is about moving in the intended direction; successfully moving around our immediate surroundings without bumping into obstructions (Montello & Sas, 2006).

Both of these parts of navigation are going to be more difficult for someone with a visual impairment (Paajala and Keranen, 2015). This project focuses on the wayfinding element of navigation (from here on the term navigation is referring to the wayfinding element only). For visually impaired people the majority of information for wayfinding is not going to be accessible (Arrasvuori and Liang, 2015; Bradley and Dunlop, 2005; Montello and Sas, 2006). Obviously visual media may not be accessible for someone with a visual impairment so other modalities need to be used; generally devices designed for visually impaired people have haptic or audio feedback (Ye et al, 2014)

According to Laakso et al (2012) navigation begins with having and understanding an internal model or cognitive map of the area that you will be navigating in. Though visually impaired people have the same cognitive spatial ability as someone with sight they have less access to the information needed to generate these maps (Golledge, 1993). Tactile maps have been used for many years to provide access to that information (Rice et al, 2013) though they tend to have limited print runs and focus on particular geographic areas (Perkins, 2002). Work has been done to automate the creation of tactile maps for example the TMAP project which generates embossed maps from GIS data (Miele et al, 2006). There has also been research using virtual maps with haptic feedback to help people to create a cognitive map (Simonnet & Vieilledent, 2011).

Once someone has a cognitive map of the area that they are going to be travelling in they need to select a route to follow. Route selection is about selecting the optimised path between the starting point and the destination (Kammoun et al, 2010). This too can also be harder for someone with a visual impairment as the shortest route might not be the safest or easiest (Dornhofer et al, 2014) (Kammoun et al, 2010). Most automatic routing software is designed for vehicle navigation (Dornhofer et al, 2014) and those that are designed for pedestrian navigation tend not to take into consideration the ability and preferences of the user (Dornhofer et al, 2014). Kammoun et al (2010) researched route selection algorithms for visually impaired people creating a system called NAVIG. Neis & Zielstra (2014) also researched finding the best route for disabled people developing an algorithm that would find the best route for the user depending on their limitations.

When a visually impaired person has a route to follow, they may need assistance to follow the route (Gaunet & Briffault, 2005). Navigational aids have been used for many years to help someone follow a route; in 1989 Brusnighan et al (1989) pioneered research into using Global Positioning Systems (GPS) to create navigational aids for visually impaired people. At this point the ideas were fairly basic with coordinates being input by the user. Today mobile phones have GPS capabilities and can be used for both vehicle and pedestrian navigation (Brock et al, 2015). Apps have been designed to help visually impaired users utilise mobile phones for navigation by using audio and haptic feedback (Williams et al, 2014) for example the Adriane GPS app has been designed for iOS devices with visually impaired people in mind. It allows the user to navigate around an area using audio and vibration and also allows the user to 'explore' an area using the screen of the device (Brock et al, 2015). Custom built GPS navigational aid devices have also been made for example Trekker Breeze by Humanware (Humanware, 2017) which has a simple interface that has been found to be easier to use (Roentgen et al, 2011).

2.3 Information requirements for visually impaired wayfinding

The above section looked at the problems that a visually impaired person might have navigating independently and some of the tools that exist to help them. For these tools to help visually impaired people they require geographic data that is relevant to someone with a visual impairment (Laakso et al, 2012). Various researchers have looked at the data requirements for visually impaired people for creating cognitive maps, route planning and for navigation.

The HaptiMap project aims to make electronic geographic information available to more people through the use of different senses like touch and hearing (Laakso et al, 2012). As part of this project research into the information needed by visually impaired people to navigate was completed and an information model was developed for content for accessible maps and location based services (Laakso et al, 2012). The data required was broken down into the following classes (with some having subclasses) footpaths, pedestrian zones, entrances, modality linkages, landmarks and obstacles. Relevant attributes for the classes was also included in the guidelines. The content guidelines can be seen in Appendix 2.

Chandler & Worsfold (2013) also researched the geographic data requirements for visually impaired people to travel more independently. They looked at the research by Laakso et al (2012) and at other research into data requirements. They interviewed visually impaired people and observed them undertaking journeys. From this research they came up with a list of data required and attributes about that data. The data requirements included designated crossing points, type of road junction/intersection, footpaths/pavements, paths attributes, accurate address locations, availability of assistive technology solutions, road type, and relevant POIs. The full table of data requirements the came up with can be seen in appendix 3.

Dornhofer et al. (2014), while looking at routing services, also looked at the different attributes that might need to be included in making a decision about routing for a visually impaired person. The main attributes that they came up with are highway, surface, footway and crossings.

This research, though categorised differently, seems to agree on the data required for navigation for example all seem to agree on the need for information about road crossings though Laakso et al include them under ‘Footpaths’. Looking at the research by Dornhofer et al. (2014) there also seems to be similar data required for route planning as to navigation. The next few paragraphs detail the elements that most of the literature seem to agree on.

Sidewalks Information about the sidewalk is needed as this is where the person will be travelling (Chandler & Worsfold, 2013). The sidewalk width and what separates it from the road can provide information to the person about which ‘walk along’ strategy to use (Gaunet & Briffault, 2005). Whether the sidewalk has any steps, and if so how many is also useful for both routing and navigation (Dornhofer et al, 2014) (Laakso et al, 2012).

Road network Information about the road network that the pedestrian is crossing or walking alongside can help to give them context. Information about the number of lanes and how busy it is can give information to the pedestrian about the risks they may face (Chandler & Worsfold, 2013).

Crossing points It can be hard for a visually impaired person to find a location to cross the road and this can be the most dangerous part of navigating (Gaunet & Briffault, 2005). Knowing the location and the type of the crossing can help a person cross the road safely (Chandler & Worsfold, 2013) (Dornhofer et al, 2014).

Transport links Visually impaired people have been shown to have problems accessing public transport (Chandler & Worsfold, 2013). Knowing where to change transport types, for example the location of a bus stop, can remove a barrier to their use (Laakso et al, 2012).

Destinations & Entrances To be able to navigate to a destination successfully it’s location is needed along with the exact location of the entrance. The entrance location is important to a visually impaired person to be able to access the destination (Laakso et al, 2012).

Obstacles Objects along the path a visually impaired person is travelling can act as either hazards to avoid or as guiding elements (Laakso et al, 2012) for example a lamp post can be a hazard as it can be in the path of a pedestrian but it can also help guide someone with limited vision.

An aerial image of Cardiff with some of these elements overlaid is shown in figure 1. This shows the amount of information needed in a small area.

The research also pointed to the need for data to have good positional accuracy to enable the user to navigate the environment (Chandler & Worsfold, 2013) (Laakso et al, 2012). This also relies on the device being used having a good degree of accuracy as well, the advice being that the accuracy should be at about 0.7m as that is the reach of a white cane (Laakso et al, 2012).

ADD IMAGE OF CARDIFF *Figure 1 - Aerial imagery of an area of Cardiff City Centre with an overlay showing the different elements that would be of use to a visually impaired person navigating. Imagery & Map Data ©2016 Google*

2.4 Geographic data sources for visually impaired people

Despite all this research into what data is required for a visually impaired person to wayfind there seems to be little research into what data meets these specifications (Chandler and Worsfold, 2013). There is a general consensus, though, that more data relevant to a visually impaired person is needed (Bradley and Dunlop, 2005; Laakso et al, 2012). In fact when Laakso et al (2012) looked at research by Roentgen et al. (2011) they commented that many of the critiques of the travel aids were actually critiques of the underlying data.

Google Maps and Apple Maps are used by many people to navigate as they are the default mapping apps on Android and Apple smartphone platforms (Samet & Fruin, 2012).] Apple Maps is one of the most used apps for navigation by visually impaired people (Chen et al, 2015) and on Android phones an app name ‘Intersection Explorer’ allows the user to build a cognitive map of the area that they are about to navigate by verbally giving information about the road layout from Google Map data (Peraković et al, 2015). Yet Chen et al (2015) found that both Google Maps and Apple Maps failed to provide some relevant information useful to a person with a visual impairment for example about when to cross the road.

Volunteered Geographic Information (VGI) uses tools to make and share, voluntarily provided, geographic information (Goodchild, 2007). This kind of data could be useful for providing data for navigational aids for visually impaired people as relevant data can be added by volunteers. The largest and most successful example of VGI is OpenStreetMap (OSM) (Mooney & Corcoran, 2011). The data in OSM is also open source which can encourage social innovation where proprietary data may not be accessible either due to cost or licensing (Arrasvuori and Liang, 2015).

OSM stores information about features in the real world as either nodes (a point with a fixed coordinate), ways (either lines or boundaries for an area) or relations (describing how individual elements work together) (Laakso et al, 2012; Ma et al, 2015). Each feature can have information about it stored in tags; tags are created by users creating a key-value pair for example 'highway=motorway'. There is no fixed classification system for tags, but there is some consensus on how they should be denoted from discussion in the OSM community (Laakso et al, 2012). The OSM wiki (OpenStreetMap, 2014) is used to record the agreed upon tags, but this is only used as a guide (Mooney & Corcoran, 2011). Data for OSM is provided by contributors uploading data from GPS devices or tracing the outline of a feature from aerial imagery (Mooney & Corcoran, 2012).

There has been some research that has used OSM data for visually impaired people. Kulyukin et al. (2008) found that OSM tends to be focussed on sighted people, though Dornhofer et al (2014) used OSM data to compare open source routing services as they thought it contained data that could be useful for visually impaired people. It was also used, along with data from FourSquare, to produce a popular iOS app for people with visual impairments called BlindSquare (Arrasvuori and Liang, 2015). Laakso et al. (2012) found no specific research into OSM data quality for visually impaired people but had the impression that it was not good.

This research shows that OSM data has great potential for use in pedestrian navigation (Laakso et al, 2012) something that could mean it would be useful for a visually impaired person. It is also often updated which could mean transitory objects are mapped which could be a barrier to navigation (Rice et al, 2013). The ability to create 'tags' means that all of the elements and attributes required for navigation could be included. Though there are concerns about the inconsistency of the tagging (Mooney and Corcoran, 2012) and the completeness of the data (Haklay, 2010). Therefore more research is required to find out whether it meets this potential to be a good data source for visually impaired people.

2.5 Data Quality

Devillers et al (2007:264) define quality as the "closeness of the agreement between data characteristics and the explicit and/or implicit needs of a user for a given application in a given area". Data quality in terms of geographic information is covered by the International Organization for Standardization (ISO) in ISO 19157 (ISO 19157:2013) which outlines six data quality elements: Completeness, Thematic Accuracy, Logical Consistency, Temporal Quality, Positional Accuracy, and Usability Element.

Completeness looks at whether features are present or not. This covers errors of omission, where there is data that is missing and errors of commission, where there is extra data that is present (Docan, 2013).

Thematic accuracy looks at the accuracy of attributes that have been assigned to a feature and the classification of the feature (INSPIRE, 2014).

Logical Consistency is concerned with not having contradictory data in a dataset e.g. only having one node in a particular location or lines intersecting at a node (Veregin, 1999)

Temporal Quality is where data is compared with changes that are made in the real world. For data for navigation this is going to equate to the data being 'up to date' (Veregin, 1999).

Positional Accuracy is about the distance between the data and the location of the feature in reality (Docan, 2013). For line data the error comes from the points which create the line and is normally measured by having a 'zone of uncertainty' around the line (Veregin, 1999).

The Usability Element looks at how the data meets a set of requirements for a particular application (Jakobsson et al, 2013).

2.6 Quality assessment of OSM data

As the biggest example of VGI (Mooney & Corcoran, 2012) the quality of the OSM dataset has interested researchers looking at geographic data quality. This research has used various methods and has looked at different aspects of data quality.

Haklay (2010) compared OSM with Ordnance Survey MasterMap and Meridian datasets. He primarily looked at streets and roads as being the ‘core feature’ collected, he also focussed on positional accuracy and completeness of the dataset. He found that while positional accuracy is good the completeness of the data is not consistent and it tends to be better in urban areas

Girres & Touya (2010) extended Haklay’s work to look at French OSM data. They also extended his research to look at the accuracy of the attributes, semantics, logical consistency, temporal accuracy, lineage and usage. They saw advantages in the flexibility of OSM and but noted that this can also limit the use cases for the data. They recommended having better defined specifications and for checking of the consistency of contributions.

The above research had focussed on the road network whereas Mooney et al (2010) looked at the similarities of shapes in OSM data for natural features including lakes and forests. They found that in urban areas there is more data. They conclude that there is a need to be able to quantify the data quality for it to be used by professionals.

Mooney and Corcoran (2012) looked at OSM attributes and found that conflicts in data are resolved quickly because of the number of contributors. They also found that there is inconsistency in the tags which can be a barrier to use, they concluded that there needs to be further work on surveying OSM contributors to look at why contributors change tags.

As there is not always a reference dataset to compare OSM data to Barron et al, (2014) researched ways of assessing data quality of OSM without using a reference dataset. The methods they used included looking at how many contributions a mapper made, how often features are updated, the logical consistency of roads and comparing attributes of similar features. These methods could assess data quality without having to find or purchase a commercial dataset. But they concluded that without a high quality reference dataset you can’t make an absolute statement about data quality.

In 2016 Senaratne et al (2016) reviewed the quality assessment methods that had been used for VGI over a number of years. They conclude that because of the diverse data characteristics and the large amount of data comparing VGI with a reference dataset is not always viable. They suggest looking at patterns in the data itself as a way of assessing data quality.

Category	Feature	Relevant.attributes	OSM.element.tag	OSM.attribute.tag
sidewalks	step	Number of steps, surface	highway=steps	step_count=, *surface=*
sidewalks	sidewalk	Location & surface	footway=, *sidewalk=*	sidewalk:left, sidewalk:both, sidewalk:right, surface=*
crossing points	crossing	Crossing type	highway=crossing, crossing=*	crossing_ref=, *traffic_signals=*, crossing=*
crossing points	underpass		tunnel=*	
transport links	bus stop		highway=bus_stop	
transport links	railway station		railway=station, building=train_station	
transport links	taxi rank		amenity=taxi	

Category	Feature	Relevant.attributes	OSM.element.tag	OSM.attribute.tag
transport links	bus station		amenity=bus_station	
destinations & entrances	building		building=replace	
destinations & entrances	entrance		entrance=*, building=entrance, door=*	
obstacles	info board		tourism=information	
obstacles	bollard		barrier=bollard	
obstacles	outdoor seating		outdoor_seating=*, amenity=biergarten	
obstacles	waste basket		amenity=waste_basket	
obstacles	bench		amenity=bench, bench=*	
obstacles	bike parking		amenity=bicycle_parking	
obstacles	call booth		amenity=telephone	
obstacles	lamp post		highway=street_lamp	
obstacles	mailbox		amenity=post_box	
obstacles	tree		natural=tree, natural=tree_row	

CHAPTER 3 - DATA AND METHODS

3.1 Introduction

This chapter will describe the data needed and the techniques that will be used to answer the aims and objectives of this research project. First it will look at the data that is going to be used and how it is going to be accessed and manipulated.

Within this research all of the six data quality elements, looked at above, will be assessed with the exception of Logical Consistency, due to it not being as relevant for this topic and usability as this will depend on all of the other elements so will be looked at in the discussion of the results. Each of the five data quality elements will be looked at in turn and the methods explained that will be used to assess whether OSM data meets that data quality element for visually impaired people. An important element that is not covered under the six data quality elements but has been looked at in previous research about data quality of VGI is that of the OSM contributors (for example Neis & Zipf (2012)) so this will also be looked at.

3.2 Data Used

OSM data is freely available under the Open Data Commons Open Database License, which means that the data can be downloaded and used freely as long as attribution is given (OpenStreetMap, 2017a). OSM data can be downloaded from a number of sources in various formats and there are a number of different tools that can be used to query or process the data (Barron et al, 2014).

There are various Application Programming Interfaces (API's) that can be used to extract OSM data. The OSM API is mainly for editing OSM data (OpenStreetMap, 2017b), it only allows downloading data for small areas so not appropriate for this project. The Overpass API is a read-only API for OSM data, large amounts of OSM data can be downloaded and queries to the API can select which elements to download by their tags, location or type of object (OpenStreetMap, 2017c). Unfortunately when trialling this method of retrieving data, queries were found to take a long time to execute and would sometime timeout.

Complete data dumps of OSM can be downloaded, either for the complete planet (OpenStreetMap, 2016d) or for particular regions (Geofabrik, 2016) & (Mapzen, 2016). Having a local copy of the data made queries quicker and allowed greater manipulation of the data. The data was downloaded in the pbf format which tends to have the smallest file sizes.

To extract the relevant entities from the data and convert it into a format that is easier to analyse two command line tools were used; osmfilter and osmconvert. Osmconvert allows the user to convert between different OSM file types e.g. pbf, osm and o5m files (OpenStreetMap, 2017d). Each file type has different advantages and disadvantages and different uses, so the ability to convert between them is useful (OpenStreetMap, 2016e). Osmfilter allows the user to filter the OSM data in various different ways including by the tags in the data (OpenStreetMap, 2017e). It was found to be a lot quicker and have more powerful filtering than the osmosis tool which is commonly used (OpenStreetMap, 2016f).

The Ordnance Survey Points of Interest dataset holds information, including the location of, businesses and education and leisure services within the UK (Ordnance Survey, 2017). The data is split into different themes, including transport and public infrastructure, and each element is classified using a classification code. This data is available from the Edina Digimap service which makes Ordnance Survey data available to be used by staff and students within UK Universities (Sutton et al, 2007).

Boundary files were downloaded from OS Boundary Line dataset (Ordnance Survey, 2017b) and the Civil Administration Area used for each study area.

To analyse the data the open source statistical software environment, R was used. R can perform statistical methods that are not available in most GIS packages and can perform spatial analysis, integrating data from multiple sources (Gómez-Rubio & López-Quílez, 2005). R can integrate with OSM data by using the OSMAR add-on (Eugster & Schlesinger, 2012).

From the analysis of previous research into information requirements for visually impaired people the following geographic features are going to be looked at in this research: sidewalks, crossing points, transport links, destinations & entrances, and obstacles. Though the road network was identified in previous research as being relevant data this research will not look at this element as it has been looked at by previous research into OSM and data quality, most notably by Haklay (2010).

For each feature type to be looked at the OSM Wiki was used to find the relevant tags to be able to query the data. Table (?) shows each feature with the OSM tag(s) used to classify that element. Also included in the table are any relevant attributes identified in the previous research and equivalent OSM tag(s).

Table (?) - Showing the features that will be studied along with the relevant OSM tags.

ADD TABLE

3.3 Methods and techniques

3.3.1 Completeness

OSM data compared with the Ordnance Survey dataset

Ordnance Survey (OS) data has been used in quality assessments of OSM data in the UK as it is considered to have high data quality as it has to conform to explicit quality standards (Veregin, 1999). Haklay (2010) uses OS data to compare OSM data with; mainly looking at completeness and positional accuracy.

To look at the completeness of OSM data for people with visual impairments, OSM data was downloaded from the Mapzen website. Osmconvert and osmfilter tools were used to extract the relevant data (the commands used can be found in appendix 7.3).

OS Points of Interest (POI) data was downloaded from Edina. The POI classification scheme (Ordnance Survey, 2013) was used to choose relevant PointX classification codes. The following element types/classification codes were found public telephones (06340460), bus stops (10590732), bus and coach stations (10570731), railway stations (10570738), taxi ranks (10570758), and subways (10550750). When studying the downloaded data it was found that the classification code for public telephones was not included in any of the data, no reason could be found for their omission. The relevant elements were extracted from the data using QGIS and filtering by the classification codes then exported as a CSV file.

Haklay (2010) found that some elements in OSM may be as far as 20 metres from the true feature location due to the way that OSM data is collected. So to compare whether a feature in the OS data was included in the OSM data the distance between the feature and OSM elements of that type were calculated. If there were any features within 20 metres then the nearest element was selected as matching the OS feature. The OSM id and the distance between the elements was then recorded for further analysis. If no elements were in 20 metres then it was recorded that no elements existed in the OSM data to match that OS feature. An R script was used to automate the process loading up the relevant OSM and OS data as spatial data frames and then looping over each OS feature (see appendix 7.2 for the script used).

OSM data compared with Ground Survey

As the OS data didn't include all of the feature types what would be relevant to a person with a visual impairment, another way of assessing the completeness of those data types was needed. From looking at other mapping data providers no datasource could be found with all of the information required, therefore a comparison with ground data was needed to check whether the relevant elements are in the OSM data. Laakso et al (2012) use a similar technique to test their AccessibilityMap data specification.

A trial in-person ground survey using a smartphone app called Epicollect+ (Aanensen et al, 2014) was done. The positional accuracy of the recorded points were found to be varied especially near tall buildings. The survey was found to be time consuming, so it was decided to complete the ground survey using a mixture of Google Street View (GSV) and Google My Maps.

Google Street View GSV has been used to carry out audits on neighbourhoods for several years (Rundle et al, 2011). GSV has also been found to have acceptable reliability when getting characteristics about routes (Vanwolleghem et al, 2014). Completing in-person ground surveys can be expensive and very time consuming meaning that they are often limited to small geographic areas (Rundle et al, 2011) so using GSV can enable more data to be collected from a wider geographic area (Rundle et al, 2011). Bader et al (2017) found that studies have shown large amounts of agreement between in-person ground surveys and surveys using GSV. There was a concern that images from GSV may be out of date but, from viewing the month/year that GSV images were captured for a number of locations in both Cardiff and Bristol, it was found that most of the images were recorded since May 2016 and no images were recorded prior to April 2014.

Google My Maps (GMM) was used to record the elements observed in GSV. GMM allows users to create custom maps and the recording of points, lines and areas over a base map (Google, no date b). The base map can be a satellite view which enables the researcher to relatively accurately compare the location to the imagery from GSV. The data from GMM can be downloaded in KML format.

As it is known that OSM completeness is not consistent (Haklay, 2010) two study cities of Bristol and Cardiff were chosen to look at areas with different data contributors. The two cities were chosen because of the researcher's knowledge of routes taken by visually impaired people, from having guided people in those cities. Using the researcher's local knowledge and the Google Maps pedestrian routing service five journeys per study area were created to locations likely to be travelled to by visually impaired people from the nearest transport hub. The number of journeys would give enough data and be able to be completed within the time constraints of the research.

Each journey was completed using GSV and the location of each feature observed that would be relevant to a visually impaired person (see table ??) was recorded in GMM with information about the type of feature it was. If there was any ambiguity about a feature the satellite imagery in GMM was used along with looking at other mapping sources. After each journey had been completed the data from GMM was downloaded in KML format.

OSM data was downloaded for the study areas from the OSM Metro Extracts website (Mapzen, 2016). The osmconvert and osmfilter tools were used to extract the relevant data in a similar process as that for the comparison with OS data.

A similar process to that used in the comparison with OS data was used to find any OSM element of the appropriate type within 20 metres of each feature and to record the OSM id and the distance to the nearest element. The R script used for this analysis can be found in appendix 7.2.

3.3.2 Thematic Accuracy

Inspection of tags of relevant OSM data features

To explore the thematic accuracy of the data the downloaded dataset from the Mapzen website was used. All of the relevant elements that the previous research found to need attribute data (see table ??) were extracted using osmconvert and osmfilter. Each element was then looked at to see whether the relevant attribute information was recorded about it. The number of elements with tags was then compared with the number of elements without tags. An R script was used to automate this process (see appendix 7.2 for the script used).

Ground survey comparison of OSM tags

The above analysis looks at whether the relevant attributes exist in the OSM data but not at whether they are accurate. No reliable reference source was found that included the detail needed to be able to compare the OSM attribute data to so a further ground survey was needed.

From the above analysis a subset of the OSM dataset was created that only included elements that had relevant attribute information. Because of the size of this dataset the subset was further reduced to only include elements within the central areas of both cities to reduce the time for the analysis.

Each element in this subset was then viewed in GSV to see if the attribute could be confirmed by viewing the imagery. For elements that could not be confirmed this way an in-person ground survey was completed. For each element it was recorded whether the attribute was accurate.

3.3.3 Temporal Quality

Inspection of the history of OSM elements

To look at temporal quality of the OSM data the version number and the timestamp of the last modification were used.

The version number starts at one when an object is created and is increased every time an object is edited (OpenStreetMap, 2017f). An object with a high version has had a number of changes to improve the quality of the element. Therefore research such as done by Keßler & de Groot (2013) have used a high version number to indicate high trustworthiness.

The timestamp for an element represents when the element was last edited (OpenStreetMap, 2017f). Neis (2014) used this information to look at the temporal quality of OSM data the assumption being that the more recent the last edit the more 'up to date' the data is. To get the version number and timestamp information the osmconvert tool was used to export this information into csv files from the Mapzen data that had previously been downloaded. This was done for the entire dataset for the study areas and for the subset of data that would be relevant to a visually impaired person. This data was imported into R software environment to be able to compare them.

3.3.4 Positional Accuracy

Comparison of OSM data with Ordnance Survey data

To look at the positional accuracy of the OSM data a comparison of the location of each element with the location of each feature in real life. The positional accuracy of the ground survey, either via GSV or in-person was not thought accurate enough for this purpose.

OS data was used by Haklay (2010) to compare with the location of OSM elements. Though many of the elements required by a visually impaired person aren't present in the OS dataset no other reference dataset was found, so this data was used for the comparison for the type of features that the OS data did include.

In the previous comparison of OSM data with OS data (see section 3.3.1) OSM data had been matched up with OS data and the distance between the feature in the OS data and the element in the OSM data recorded. R was then used to summarise information about the distance between OSM elements and features in real life for different feature types.

3.3.5 Contributors

Analysis of OSM contributors

Senaratne et al (2017), in their review of methods of assessing data quality for VGI, extend the six data quality indicators to include some more abstract indicators namely trustworthiness, credibility, text content quality, vagueness, local knowledge, experience, recognition, and reputation. Most of these indicators have some relation to the contributor of the data so obtaining knowledge about the OSM contributors is important when looking at data quality.

Senaratne et al (2017) say that contributor experience can be measured by the number of contributions that the user has made and the time that they have been registered as a contributor. To look at this for the data relevant to a visually impaired person osmconvert was used to extract the uid field as a csv file for both the relevant data and for the entire dataset in the study areas. This data was then imported into R to summarise the number of contributions by each contributor in the area. Information about the length of time that each contributor has been editing data was added to this by using the OSM comments API (https://osm-comments-api.mapbox.com/).

Senaratne et al (2017) also use local knowledge as a quality indicator. As OSM does not provide data about the home location of the user (Neis & Zipf, 2012) the number of contributions within the study boundary will be compared with their total number of contributions overall. The assumption is that if a large percentage of a contributors activity is within the study boundary then they are likely to have a good local knowledge. To get information about the total number of edits for each contributor the OSM comments API was used. This information was then used to split the contributors into three groups as defined by Neis and Zipf (2012): - Senior Mappers with 1,000 or more edits - Junior Mappers with at least 10 but less than 1,000 edits - Nonrecurring Mappers with less than 10 edits It was also considered that the motivations of the user could give information about the data quality. Though the motivations of every contributor can't be assessed directly the 'description' section of the OSM contributors profile can give an indication of the the motivations. The OSM API was used to extract this data for each of the users who provided relevant data. A text search was then done for keywords to find out information about the user for example a contributor with the word 'Bristol' in their profile is likely to be based in the area of Bristol and mainly contributing data for that area.

Survey of OSM contributors

Budhathoki and Haythornthwaite (2012) looked at the motivations of people contributing to OSM by using an online survey. To look at the motivations of people contributing data relevant for visually impaired people and to find out what knowledge contributors had about data relevant to visually impaired people a survey was completed. The survey was distributed through several OSM mailing lists both mailing lists specifically about disability and more general mailing lists. This was done to get information and comments from both people who have some awareness of the needs of visually impaired people and those who don't. The survey was conducted online using Google Forms (Google ,no date a)

The survey asked questions about the person's knowledge of OSM data elements that are relevant to a person with a visual impairment, the reasons that they contribute data in general, the reasons that they contribute data that is relevant to a person with a visual impairment (if relevant) and about ways that could encourage more people to contribute relevant data.