Haklay, M., 2008, Please note that this is an early version of this article, and that the definitive, peer-reviewed and edited version of this article is published in *Environment and Planning B: Planning and Design*. Please refer to the publication website for the definitive version.

How good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey datasets

Abstract

Within the framework of Web Mapping 2.0 applications, the most striking example of a geographical application is the OpenStreetMap project. OpenStreetMap aims to create a free digital map of the world and is implemented through the engagement of participants in a mode similar to software development in Open Source projects. The information is collected by many participants, collated on a central database and distributed in multiple digital formats through the World Wide Web (Web). This type of information was termed 'Volunteered Geographical Information' (VGI) by Mike Goodchild (2007). However, to date there has been no systematic analysis of the quality of VGI. This paper aims to fill this gap by analysing OpenStreetMap information. The examination starts with the characteristics of OpenStreetMap contributors, followed by analysis of its quality through a comparison with Ordnance Survey datasets. The analysis focuses on London and England, since OpenStreetMap started in London in August 2004 and therefore the study of these geographies provides the best understanding of the achievements and difficulties of VGI.

The analysis shows that OpenStreetMap information can be fairly accurate: on average within about 6 metres of the position recorded by the Ordnance Survey, and with approximately 80% overlap of motorway objects between the two datasets. In the space of four years, OpenStreetMap has captured about 29% of the area of England, of which approximately 4% are digitised lines without a complete set of attributes. The paper concludes with a discussion of the implications of the findings to the study of VGI as well as suggesting future research directions.

1. Introduction

While the use of the Internet and the World Wide Web (Web) for mapping applications is well into its second decade, the landscape has changed dramatically since 2005 (Haklay *et al.*, 2008). One expression of this change is the emerging neologism that follows the rapid technological developments. While terms such as Neogeography, Mapping Mash-ups, Geotagging and Geostack may seem alien to veterans in the area of Geographical Information Systems (GIS), they can be mapped to existing terms that have been in use for decades; so Mash-up is a form of interoperability between geographical databases, Geotagging means a specific form of Georeferencing or Geocoding, the Geostack is a GIS and Neogeography is the sum of these terms in an attempt to divorce the past and conquer new (cyber)space. Therefore, the neologism does not represent new ideas, rather a zeitgeist which is indicative of the change that has happened.

Yet, it is hard not to notice the whole range of new websites and communities – from the commercial Google Maps to the grassroots OpenStreetMap, and to applications such as Platial – that have emerged. The sheer scale of new mapping applications is evidence of a step change in the Geographical Web (GeoWeb). Mapping has gained prominence within the range of applications known as Web 2.0, and the attention that is given to this type of application in the higher echelons of the hi-tech circles is exemplified by a series of conferences, 'Where 2.0', which were started in 2006 by O'Reilly Media –

probably the leading promoters of hi-tech knowhow: 'GIS has been around for decades, but is no longer only the realm of specialists. The Web is now flush with geographic data, being harnessed in a usable and searchable format.' (Where 2.0, 2008)

While Haklay, Singleton and Parker (2008) and Elwood (2009) provide an overview of this Web Mapping 2.0 landscape, one of the most interesting aspects is the emergence of crowdsourced information. Crowdsourcing is one of the most significant and potentially controversial developments in Web 2.0. This term developed from the concept of outsourcing where business operations are transferred to remote cheaper locations (Friedman, 2006). Similarly, crowdsourcing is how large groups of users can perform functions that are either difficult to automate or expensive to implement (Howe, 2006).

The reason for the controversial potential of crowdsourcing is that it can be a highly exploitative activity, in which participants are encouraged to contribute to an alleged greater good when, in reality, the whole activity is contributing to an exclusive enterprise that profits from it. In such situations, crowdsourcing is the ultimate cost reduction for the enterprise, in which labour is used without any compensation or obligation between the employer and the employee.

On the other hand, crowdsourcing can be used in large-scale community activities that were difficult to implement and maintain before Web 2.0. Such community activities can be focused on the development of software, or more recently on the collection and sharing of information. This 'Commons-Based Peer Production' has attracted significant attention (Benkler and Nissenbaum, 2006). Tapscott and Williams (2006) note, in relation to this form of activity, that 'in many peer production communities, productive activities are voluntary and non-monetary'. In these cases, the participants contribute to achieve a useful goal that will serve their community, and very frequently the wider public. This is well exemplified by the creation of Open Source software such as the Firefox web browser or the Apache web server: both are used by millions, while being developed by a few hundred people. Even in cases where the technological barrier for participation is not as high as in software development, the number of participants is much smaller than the users. For example, in Wikipedia, well over 99.8% of visitors to the site do not contribute anything (Nielsen, 2006), yet this does not deter the contributors – on the contrary, they gain gratification from the usefulness of their contributions.

The use of large-scale crowdsourcing activities to create reliable sources of information or high-quality software is not without difficulties. These activities – especially the commons-based one – are carried out by a large group of volunteers, who work independently and without much co-ordination, each concentrating on their own interests. In successful commons-based peer-production networks, there are lengthy deliberations within the communities about the directions that the project should take or how to implement a specific issue. Even after such deliberation, these projects have a limited ability to force participants to a specific course of action, other than banish them from the project at the cost of losing a contributor (and usually a significant one). Furthermore, and especially in information-based activities, the participants are not professionals but amateurs (Keen, 2007) and therefore do not follow common standards in terms of data collection, verification and use. This is a core issue which is very frequently discussed and debated in the context of crowdsourcing activities (Tapscott and Williams, 2006; Friedman, 2006).

The potential of crowdsourced geographical information has captured the attention of researchers in GIS (including Goodchild, 2007a, 2007b, 2008; Sui, 2008; Elwood, 2009). Goodchild has coined a term to describe this activity as 'Volunteered Geographical Information' (VGI). In the area of Geographical Information, the question of information quality has been at the centre of the research agenda since the first definition of GIScience (Goodchild, 1992). Therefore, in light of the data collection by amateurs, the distributed nature of the data collection and the loose co-ordination in terms of standards, one of the

significant core questions about VGI is 'how good is the quality of the information that is collected through such activities?' This is a crucial question about the efficacy of VGI activities and the value of the outputs for a range of applications, from basic navigation to more sophisticated applications such as site location planning.

To answer this question, the OpenStreetMap (OSM) project provides a suitable case study. OSM aims to create map data that are free to use, editable and licensed under new copyright schemes – such as Creative Commons – which protect the project from unwarranted use by either participants or a third party (Benkler and Nissenbaum, 2006). A key motivation for this project is to enable free access to current digital geographical information across the world, as such information is not available until now. In many western countries this information is available from commercial providers and national mapping agencies, but it is considered to be expensive and is out of reach of individuals and grass-roots organisations. Even in the US, where basic road information is available through the US Census Bureau TIGER/Line programme, the details that are provided are limited (streets and roads only) and do not include green space, landmarks and the like. Also, due to budget limitations, the update cycle is slow and does not take into account rapid changes. Thus, even in the US, there is a need for detailed free geographical information.

OSM information can be edited online through a wiki-like interface where, once a user has created an account, the underlying map data can be viewed and edited. In addition to this lightweight editing software package working within the browser, there is a stand-alone editing tool, more akin to a GIS package. A number of sources have been used to create these maps including uploaded Global Positioning System (GPS) tracks, out-of-copyright maps and Yahoo! aerial imagery which was made available through collaboration with this search engine. Unlike Wikipedia, where the majority of content is created at disparate locations, the OSM community also organises a series of local workshops (called 'mapping parties'), which aim to create and annotate content for localised geographical areas (see Perkins and Dodge, 2008). These events are designed to introduce new contributors to the community with hands-on experience of collecting data, while positively contributing to the project overall by generating new information and street labelling as part of the exercise. The OSM data are stored on servers at University College London, and Bytemark, which contributes the bandwidth for this project. Whilst over 50,000 people have contributed to the map as of August 2008, it is a core group of about 40 volunteers who dedicate their time to create the technical infrastructure for a viable data collection and dissemination service. This includes the maintenance of the servers, writing the core software that handles the transactions with the server when adding and editing geographical information, and creating cartographical outputs. For a detailed discussion of the technical side of the project, see Haklay and Weber (2008).

With OSM, it is possible to answer the question of VGI quality by comparing the dataset against Ordnance Survey (OS) datasets in the UK. As OSM started in London, and thus the city represents the place that received the longest ongoing attention from OSM participants, it stands to reason that an examination of the city and of England will provide an early indication about the quality of VGI.

This paper discusses an analysis of the quality of the OSM dataset, evaluating its positional and attribute accuracy, completeness and consistency. In light of this analysis, the paper suggests the fitness for purpose of OSM information and some possible directions for future developments. However, before turning to the analysis, a short discussion of evaluations of geographical information quality will help to set the scene.

2. How to evaluate the quality of geographical information

The problem of understanding the quality of geographical databases was identified many years ago, and received attention from surveyors, cartographers and geographers (van Oort, 2006). Van Oort identified work on the quality of geographical information dating back to the late 1960s and early 1970s.

With the emergence of Geographical Information Systems in the 1980s, this area of research experienced rapid growth, receiving attention from leading figures in the area of Geographical Information Science including Peter Burrough and Andrew Frank (1996), Mike Goodchild (1995), Peter Fisher (1999), Nick Chrisman (1984) and many others (see van Oort, 2006, for a comprehensive review of the area). By 2002, quality aspects of geographical information had been enshrined in the International Organisation for Standards (ISO) codes 19113 (Quality principles) and 19114 (Quality evaluation procedures) under the aegis of Technical Committee 211. In their review of these standards, Kresse and Fadaie (2004) identified the following aspects of quality: completeness, logical consistency, positional accuracy, temporal accuracy, thematic accuracy, purpose, usage and lineage.

Van Oort’s (2006) synthesis of various quality standards and definitions is more comprehensive and identifies the following aspects:

Lineage – this aspect of quality is about the history of the dataset, how it was collected and evolved.

Positional accuracy – this is probably the most obvious aspect of quality and evaluates how well the coordinate value of an object in the database relates to the reality on the ground.

Attribute accuracy – as objects in a geographical database are represented not only by their geometrical shape but also by additional attributes, this measure evaluates how correct these values are.

Logical consistency – this is an aspect of the internal consistency of the dataset, in terms of topological correctness and the relationships that are encoded in the database.

Completeness – this is a measure of the lack of data; that is, an assessment of how many objects are expected to be found in the database but are missing as well as an assessment of excess data that should not be included. In other words, how comprehensive the coverage of real-world objects is.

Semantic accuracy – this measure links the way in which the object is captured and represented in the database to its meaning and the way in which it should be interpreted.

Usage, purpose and constraints – this is a fitness-for-purpose declaration that should help potential users in deciding how the data should be used.

Temporal quality – this is a measure of the validity of changes in the database in relation to real-world changes and also the rate of updates.

Naturally, the definitions above are shorthand and aim to explain the principles of geographical information quality. The burgeoning literature on geographical information quality provides more detailed definitions and discussion of these aspects (see van Oort, 2006; Kresse and Fadaie, 2004).

To understand the amount of work that is required to achieve a high-quality geographical database, the Ordnance Survey provides a good example of monitoring completeness and temporal quality (based on Cross *et al.*, 2005). To achieve this goal, the Ordnance Survey has an internal quality assurance process known as ‘The Agency Performance Monitor’. This is set by the UK government and requires that ‘some 99.6% significant real-world features are represented in the database within six months of completion’. Internally to Ordnance Survey, the operational instruction based on this criterion is the maintenance of the Ordnance Survey large-scale database currency at an average of no more than 0.7 House Units of

unsurveyed major change, over six months old, per Digital Map Unit (DMU). DMUs are inherently map tiles, while House Units are a measure of data capture, with the physical capture of one building as the basic unit. To verify that they comply with the criterion, every six months the Ordnance Survey analyses the result of auditing over 4000 DMUs, selected through stratified sampling, for missing detail by sending semi-trained surveyors with printed maps on the ground. This is a significant and costly undertaking but it is an unavoidable part of creating a reliable and authoritative geographical database. Noteworthy is that this work focuses on completeness and temporal quality, while positional accuracy is evaluated through a separate process.

As this type of evaluation is not feasible for OSM, a desk-based approach was taken using two geographical datasets: the Ordnance Survey dataset and OSM dataset. The assumption is that, at this stage of OSM development, the Ordnance Survey dataset represents higher accuracy and overall quality (at least positional and attribute). Considering the lineage and investment in the Ordnance Survey dataset, this should not be a contested statement. This type of comparison is common in spatial data quality research (see Hunter, 1999; Goodchild *et al.,* 1992).

3. Datasets used and comparison framework

A basic problem that is inherent in a desk-based quality assessment of any spatial dataset is the selection of the comparison dataset. The explicit assumption in any selection is that the comparison dataset is of higher quality and represents a version of reality that is consistent in terms of quality, and is therefore usable to expose shortcomings in the dataset that is the subject of the investigation.

Therefore, a meaningful comparison of OSM information should take into account the characteristics of this dataset. Due to the dataset collection method, the OSM dataset cannot be more accurate than the quality of the GPS receiver (which usually captures a location within 6-10 metres) and the Yahoo! imagery, which outside London provides about 15m resolution. This means that we can expect the OSM dataset to be within a region of about 20m from the true location under ideal conditions. Therefore, we should treat it as a generalised dataset. Furthermore, for the purpose of the comparison, only streets and roads will be used, as they are the core feature that is being collected by OSM volunteers.

Based on these characteristics, Navteq or TeleAtlas datasets, where comprehensive street level information without generalisation is available, should be the most suitable. They are collected under standard processes and quality assurance procedures, with a global coverage. Yet, these two datasets are outside the reach of researchers without incurring high costs of purchasing the data, and a request to access such a dataset for the purpose of comparing it to OSM was turned down.

As an alternative, the Ordnance Survey datasets were considered. Significantly, the Ordnance Survey willingly provided their datasets for this comparison. Of the range of Ordnance Survey vector datasets, Meridian 2 (for the sake of simplicity, 'Meridian') and MasterMap were used. Meridian is a generalised dataset and, due to reasons that are explained below, it holds some characteristics that make it similar to OSM and suitable for comparison. The MasterMap Integrated Transport Layer (ITN) dataset is a large-scale dataset with high accuracy level but, due to data volumes, it can be used only in several small areas for a comprehensive comparison.

As Meridian is central to the comparison, it is worth paying attention to its characteristics. Meridian is a vector dataset that provides coverage of Great Britain with complete details of the national road network: 'Motorways, major and minor roads are represented in the dataset. Complex junctions are collapsed to single nodes and multi-carriageways to single links. To avoid congestion, some minor roads and cul-de-sacs less than 200m are not represented ... Private roads and tracks are not included.' (OS,

2007, p. 24.) The source of the road network is high-resolution mapping (1:1250 in urban areas, 1:2500 in rural areas and 1:10,000 in moorland).

Meridian is constructed so that the node points are kept in their original position while, through a process of generalisation, the road centreline is filtered to within a 20m region of the original location. The generalisation process decreases the number of nodes to reduce clutter and complexity. Thus, Meridian's position accuracy is 5 metres or better for the nodes, and within 20 metres of the real-world position for the links between the nodes.

The Ordnance Survey describes Meridian as a dataset suitable for applications from environmental analysis to design and management of distribution networks for warehouses to health planning.

Two other sources were used to complete the comparison. First, the 1:10,000 raster files from the OS. This is the largest scale raster product that is available from the Ordnance Survey. These are based on detailed mapping, and went through a process of generalisation that leaves most of the features intact. It is a highly detailed map, and thus suitable for locating attribute information and details of streets and other features that are expected to be found in OSM.

The second source is the Lower Level of Super Output Areas (SOAs), which is provided by the Ordnance Survey and the Office of National Statistics and is based on the Census. SOAs are about the size of a neighbourhood and are created through a computational process by merging the basic Census units. This dataset was combined with the Index of Deprivation 2007 (ID 2007), created by the Department of Communities and Local Government and which indicates the socio-economic status of each SOA. This dataset is used in section 4.5 for the analysis of the socio-economic bias of VGI.

The OSM dataset that was used in this comparison was from the end of March 2008, and was based on roads information created by Frederik Ramm and available on his website Geofabrik. The dataset is provided as a set of thematic layers (building, natural, points, railways, roads and waterways), which are classified according to their OSM tags.

The process of comparison started from an evaluation of positional accuracy, first by analysing motorways, A-roads and B-roads objects in the London area, and then by closely inspecting five, randomly selected, Ordnance Survey tiles at 1:10,000 resolution, covering 113 square kilometres. After this comparison, an analysis of completeness was carried out: first through a statistical analysis across England, followed by a detailed visual inspection of the 1:10,000 tiles. Finally, statistical analysis of SOAs and ID 2007 was carried out.

4. Detailed methodology and results

For this preliminary study, two elements of the possible range of quality measures were reviewed – positional accuracy and completeness. Firstly, positional accuracy is 'the best established issue of accuracy in the mapping science' (Chrisman, 1991) and therefore must be tested. Positional accuracy is significant in the evaluation of fitness for use of data that was not created by professionals and was without stringent data collection standards. Secondly, completeness is significant in the case of VGI, as data collection is done by volunteers who are collecting information of their own accord without top-down co-ordination that ensures systematic coverage. At this stage of the development of VGI, the main question is the ability of these loosely organised peer-production collectives to cover significant areas in a way that renders their dataset useful.

4.1 Positional accuracy: motorways, A-roads and B-roads comparison[1]

The evaluation of the positional accuracy of OSM can be carried out against Meridian, since the nodes of Meridian are derived from the high-resolution topographical dataset and thus are highly accurate. However, the fact that the number of nodes has been diluted by the application of a filter and the differing digitising methods means that the two datasets have a different number of nodes. Furthermore, in the case of motorways, OSM represents these as a line object for each direction, whereas Meridian represents them as a single line. This means that matching on a point-by-point basis would be inappropriate in this case.

Motorways were selected as the objects for comparison as they are significant objects on the landscape so the comparison will evaluate the data capture along lengthy objects, which should be captured in a consistent way. In addition, at the time of the comparison, motorways were completely covered by the OSM dataset, so the evaluation does not encounter completeness problems. The methodology used to evaluate the positional accuracy of motorway objects across the two datasets was based on Goodchild and Hunter (1997) and Hunter (1999). The comparison is carried out by using buffers to determine the percentage of line from one dataset that is within a certain distance of the same feature in another dataset of higher accuracy (Figure 1).

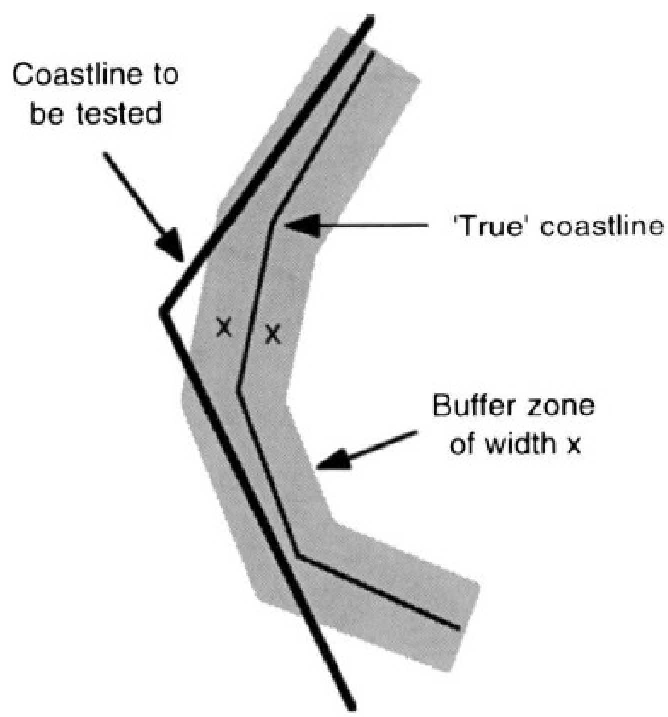


Figure 1 – Goodchild and Hunter buffer comparison method. The buffer of width x is created around the high-quality object, and the percentage of the tested object that falls within the buffer is evaluated (Source: Goodchild & Hunter, 1997).

[1] This section is based on the M.Eng. reports of Naureen Zulfiqar and Aamer Ather

The preparation of the datasets for comparison included some manipulation. The comparison was carried out for the motorways in the London area on both datasets to ensure that they represent roughly the same area and length. Complex slip road configurations were edited in the OSM dataset to ensure that the representation was similar to the one in Meridian. The rest of the analysis was carried out by creating a buffer around each dataset, and then evaluating the overlap. As the Ordnance Survey represents the two directions as a single line, it was decided the buffer around the Meridian should be set at 20 metres (as this is the filter that the Ordnance Survey applies in the creation of the line) and, to follow Goodchild and Hunter's method, the OSM dataset was buffered with a 1-metre buffer to calculate the overlap.

The results are displayed in Table 1.

Motorway	Percentage
M1	87.36%
M2	59.81%
M3	71.40%
M4	84.09%
M4 Spur	88.77%
M10	64.05%
M11	84.38%
M20	87.18%
M23	88.78%
M25	88.80%
M26	83.37%
M40	72.78%
A1(M)	85.70%
A308(M)	78.27%
A329(M)	72.11%
A404	76.65%

Table 1 – Percentage overlap between Meridian and OSM buffers Based on this analysis, we can conclude that with an average overlap of nearly 80% and variability from 60% up to 89%, the OSM dataset provides a good representation of motorways.

A further analysis was carried out using five tiles (5km x 5km) of Ordnance Survey MasterMap ITN, randomly selected from the London area, to provide an estimation of the accuracy of capture of A-roads and B-roads, which are the smaller roads in the UK hierarchy. For this analysis, the buffer that was used for A-roads was 5.6m, and for B-roads 3.75m. Thus, this test was using a higher accuracy dataset (MasterMap) and stringent comparison conditions in the buffers. This comparison included over 246km of A-Roads and the average overlap between MasterMap and OSM was 88%, with variability from 21% to 100%. In the same areas there were 68km of B-roads, which were captured with an overall overlap of 77% and variability from 5% to 100%. The results of this comparison are presented in Figure 2.

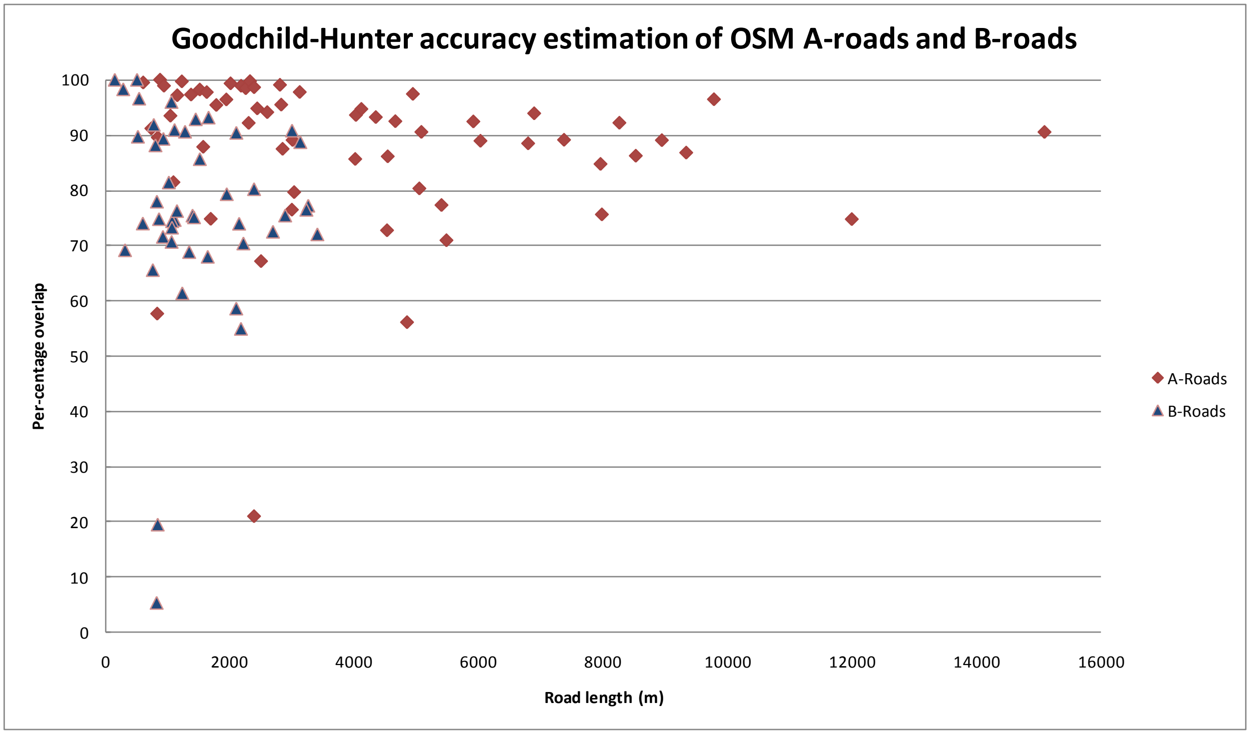


Figure 2 – Comparison of A- and B-roads with Ordnance Survey MasterMap.

4.2 Positional accuracy: urban areas in London

In addition to the statistical comparison, a more detailed, visual comparison was carried out across 113 square kilometres in London using five Ordnance Survey 1:10,000 raster tiles (TQ37ne – New Cross, TQ28ne – Highgate, TQ29nw – Barnet, TQ26se – Sutton, and TQ36nw – South Norwood). The tiles were randomly selected from the London area. In each one of them, the tiles were inspected visually and 100 samples were taken to evaluate the difference between the Ordnance Survey centreline and the location that is recorded in OSM.

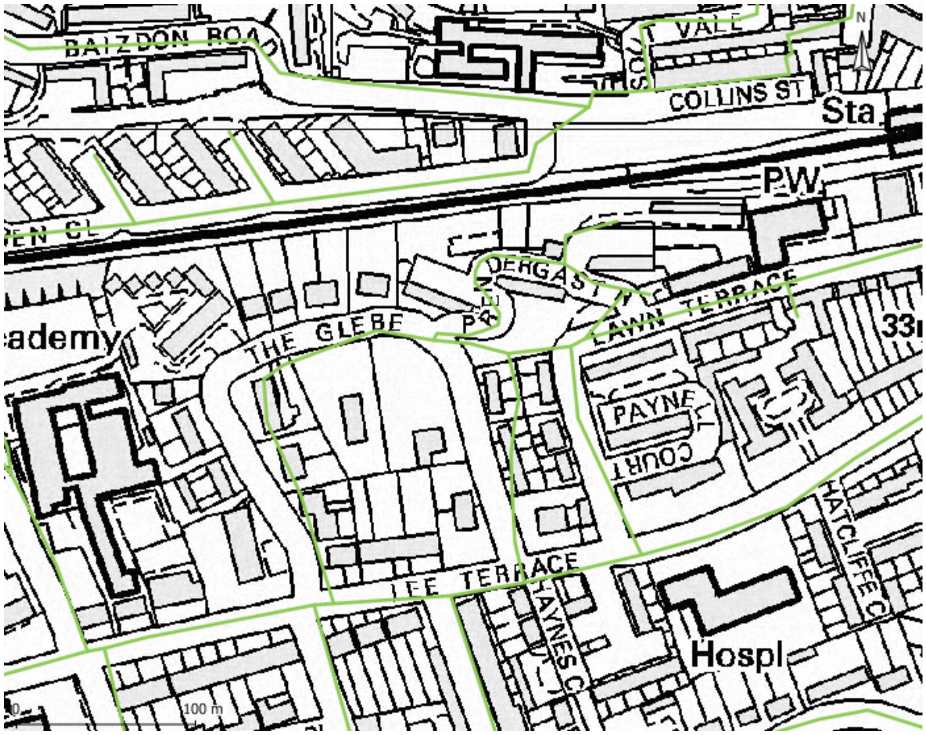
The average differences between the Ordnance Survey location and OSM are provided in Table 2.

Area	Average difference (m)
Barnet	6.77
Highgate	8.33
New Cross	6.04
South Norwood	3.17
Sutton	4.83
Total	*5.83*

Table 2 – Positional accuracy across five areas of London

Notice the difference in averages between the areas. In terms of the underlying measurements, in the best areas many of the locations are within a metre or two of the location, whereas in Barnet and Highgate distances of up to 20 metres from the Ordnance Survey centreline were recorded. Figure 3

provides examples from New Cross (A), Barnet (B) and Highgate(C), which show the overlap and mismatch between the two datasets.



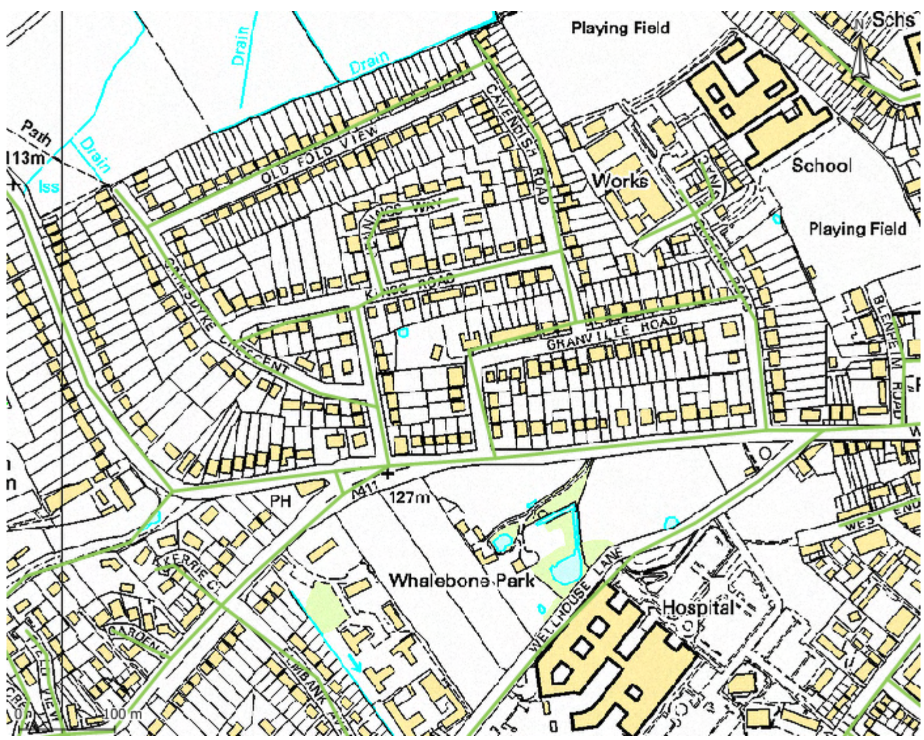


Figure 3 – Examples of overlapping OSM and Ordnance Survey maps for New Cross (A), Barnet (B) and Highgate (C). The lighter lines that overlay the map are OSM features.

The visual examination of the various tiles shows that the accuracy and attention to detail differs between areas. This can be attributed to digitisation, data collection skills and the patience of the person who carried out the work.

4.3 Completeness: length comparison

After gauging the level of positional accuracy of the OSM dataset, the next issue is the level of completeness. While Steve Coast, the founder of OSM, stated 'it's important to let go of the concept of completeness' (GISPro, 2007, p.22), it is important to know which areas are well covered and which are not – otherwise, the data can be assumed to be unusable. Furthermore, the analysis of completeness can reveal other characteristics that are relevant to other VGI projects.

Here, the difference in data capture between OSM and Meridian provided the core principle for the assessment of comparison:

Because Meridian is generalised, excludes some of the minor roads, and does not include foot and cycle paths, in every area where OSM has a good coverage, the total length of OSM roads must be longer than the total length of Meridian features.

This aspect can be compared across the whole area of Great Britain, but as OSM started in England (and more specifically in London) a comparison across England was more appropriate and manageable.

To prepare the dataset for comparison, a grid at a resolution of 1km was created across England. Next, as the comparison is trying to find the difference between OSM and Meridian objects, and to avoid the inclusion of coastline objects and small slivers of grid cells, all incomplete cells with an area less than a square kilometre were eliminated. This meant that out of the total area of England of 132,929 sq km, the comparison was carried out on 123,714 sq km (about 93% of the total area).

The first step was to project the OSM dataset onto the British National Grid, to bring it to the same projection as Meridian. The grid was then used to clip all the road objects from OSM and from Meridian in such a way that they were segmented along the grid lines. This step enabled the comparison of the two sets in each cell grid across England. For each cell, the following formula was calculated:

$$\sum(\text{OSM roads length}) - \sum(\text{Meridian roads length})$$

The rest of the analysis was carried out through SQL queries, which added up the length of lines that were contained in or intersected the grid cells. The clipping process was carried out in MapInfo, whereas the analysis was in Manifold GIS.

The results of the analysis show the current state of OSM completeness. At the macro level, the total length of Meridian roads is 302,349,778 metres, while OSM is 209,755,703 metres. Thus, even at the highest level, the OSM dataset total length is 69% of Meridian. It is important to remember that, in this and in the following comparisons, Meridian is an incomplete and generalised coverage, and thus this is an underestimation of the total length of roads for England. Yet, considering the fact that OSM has been around for a mere four years, this is a significant and impressive rate of data collection.

There are 16,300 sq km in which neither OSM nor Meridian has any feature. Out of the remainder, in 70.7% of the area, Meridian provides a better, more comprehensive coverage than OSM. In other words OSM volunteers have provided an adequate coverage for 29.3% of the area of England in which we should expect to find features.

Empty cells	16,300 (13.2%)
Meridian 2 more detailed than OSM	75,977 (61.4%)
OSM more detailed than Meridian 2	31,437 (25.4%)
Total	**123,714**

Table 3 – Length comparison: OSM and Meridian 2 in sq km

Naturally, the real interest lies in the geography of these differences. The centres of the big cities of England (such as London, Manchester, Birmingham, Newcastle and Liverpool) are well mapped using this measure. However, in the suburban areas, and especially in the boundary between the city and the rural area that surrounds it, the quality of coverage drops very fast and there are many areas that are not covered very well. Figure 4 shows the pattern across England. The areas that are marked in black show where OSM is likely to be complete, while the grey indicates incompleteness. The white areas are the locations where there is no difference between the two (where the difference between the two datasets is up to +/-1m).

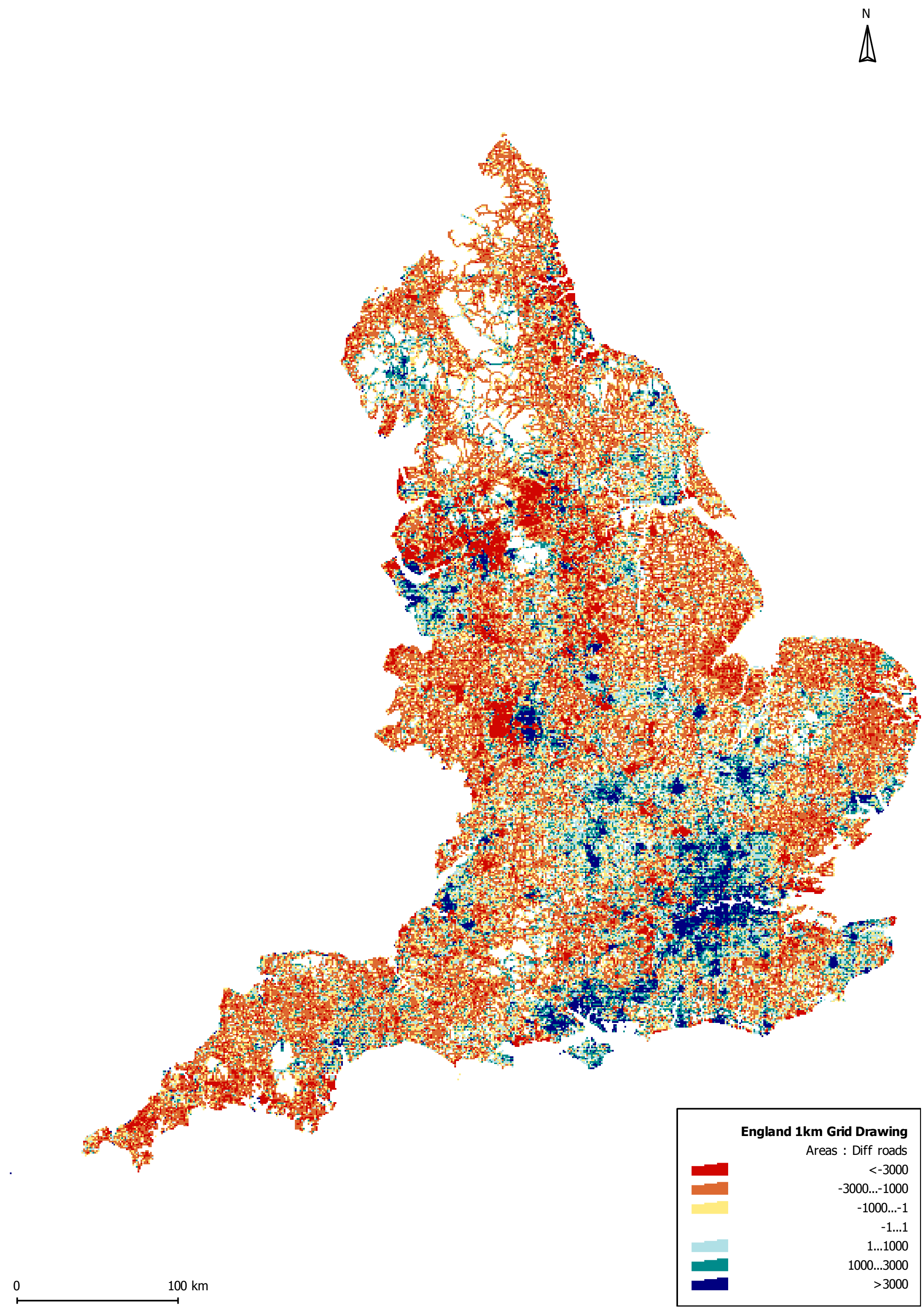


Figure 4 – Length difference between OSM and Meridian datasets. Areas of good OSM coverage are in black, and where it is lacking are in grey.

Figure 4 highlights the importance of the Yahoo! imagery – the rectangular area around London where high-resolution imagery is available, and therefore data capture is easier, is clearly visible.

Following this comparison, which inherently compares all the line features that are captured in OSM, including footpaths and other minor roads, a more detailed comparison was carried out. This time, only OSM features that were comparable to the Meridian dataset were included (e.g. motorway, major road, residential).

Noteworthy is that this comparison moves into the area of attribute quality, as a road that is included in the database but without any tag will be excluded. Furthermore, the hypothesis that was noted above still stands – in any location in which the OSM dataset has been captured completely, the length of OSM objects must be longer than Meridian objects.

Empty cells[2]	17,632 (14.3%)
Meridian 2 more detailed than OSM	80,041 (64.7%)
OSM more detailed than Meridian 2	26,041 (21.0%)
Total	123,714

Table 4 – Length comparison with attributes: OSM and Meridian 2 in sq km

Under this comparison, the OSM dataset is providing coverage for 24.5% out of the total area that is covered by Meridian. Figure 5 provides an overview of the difference in the London area.

[2] The rise in the number of empty cells is due to the removal of cells that contain OSM information on footpaths and similar features.

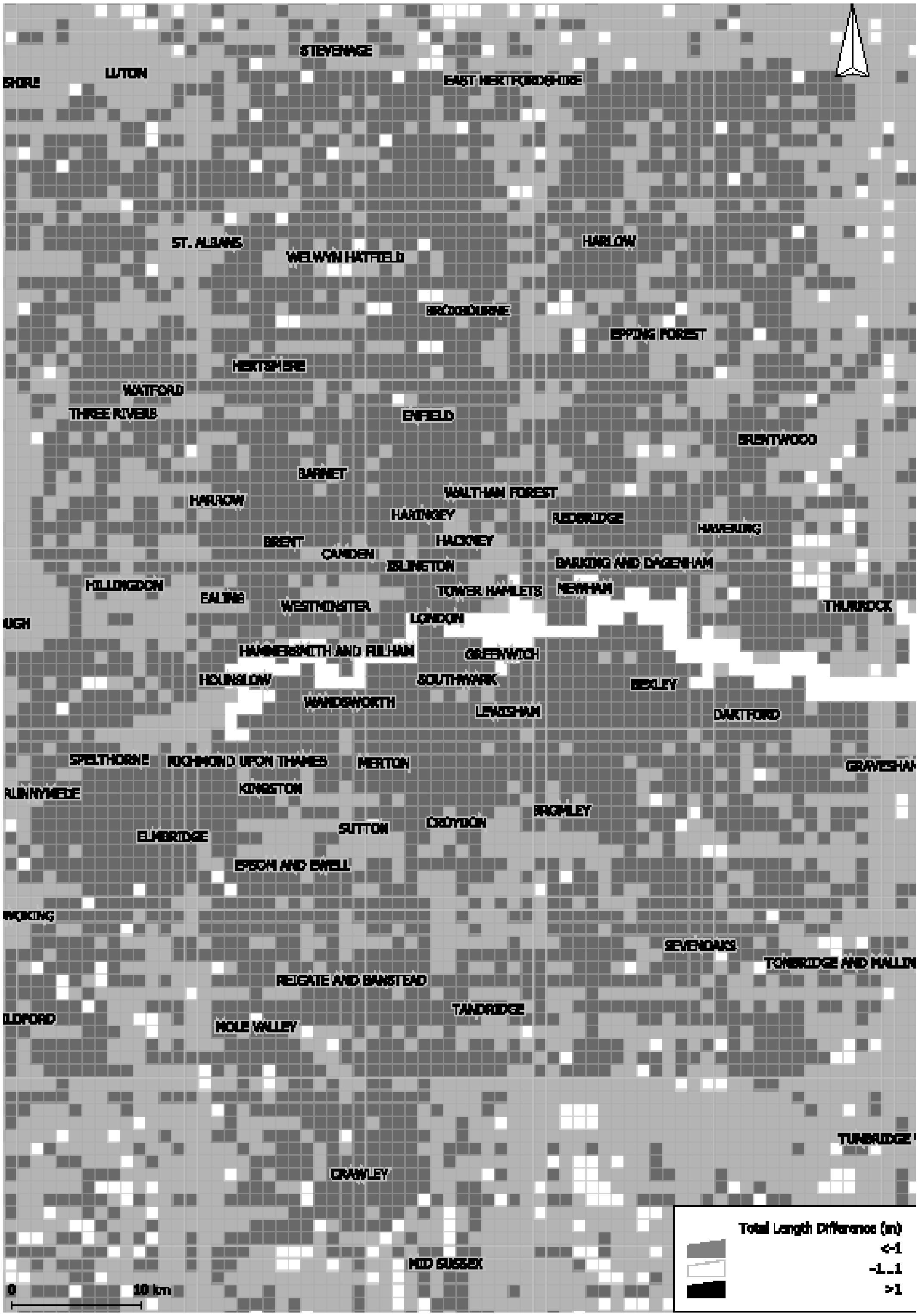


Figure 5 – difference between Ordnance Survey and OSM, including attributes indicating comparable categories to Meridian.

4.4 Completeness: urban areas in London

Another way to evaluate the completeness of the dataset is by visual inspection of the dataset against another dataset. Similar to the method that was described above for the detailed analysis of urban areas, 113 square kilometres in London were examined visually to understand the nature of the incompleteness in OSM. The five 1:10,000 raster tiles are shown in Figure 6, and provide a good cross-section of London from the centre to the edge. Each red circle on the image indicates an omission of a detail or a major mistake in digitising (such as a road that passes through the centre of a built-up area).

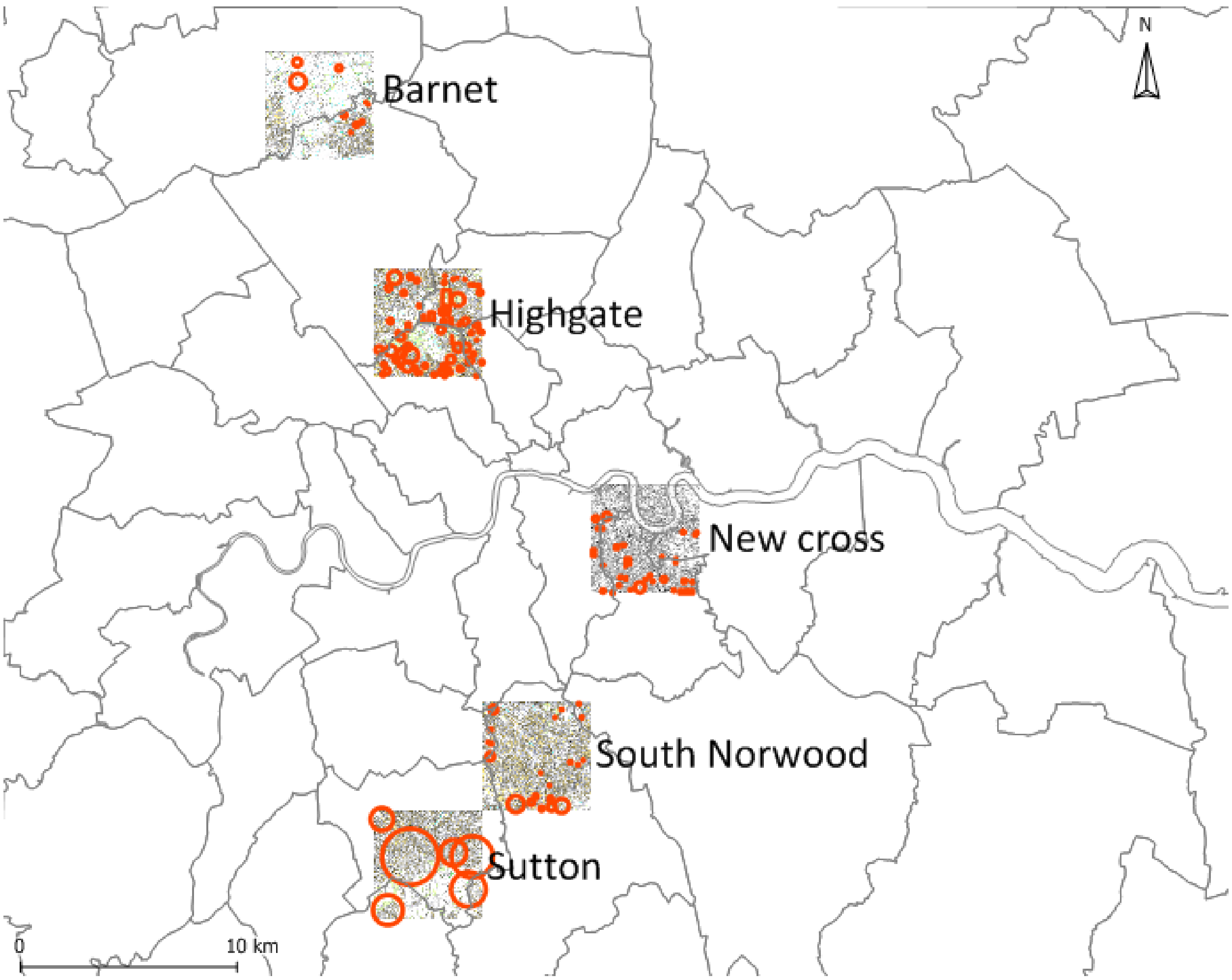


Figure 6 – Overview of completeness across five areas in London.

Out of the five tiles, two stand out dramatically. The Highgate tile includes many omissions, and, as noted in the previous section, also exemplifies sloppy digitisation, which impact the positional accuracy of the dataset. As Figure 7 shows, open spaces are missing, as well as minor roads. Notice that some of OSM lines are at the edge of the roads and some errors in digitising can be identified clearly. The Sutton tile contains large areas that are completely missing – notice the size of the circles on Figure 6.

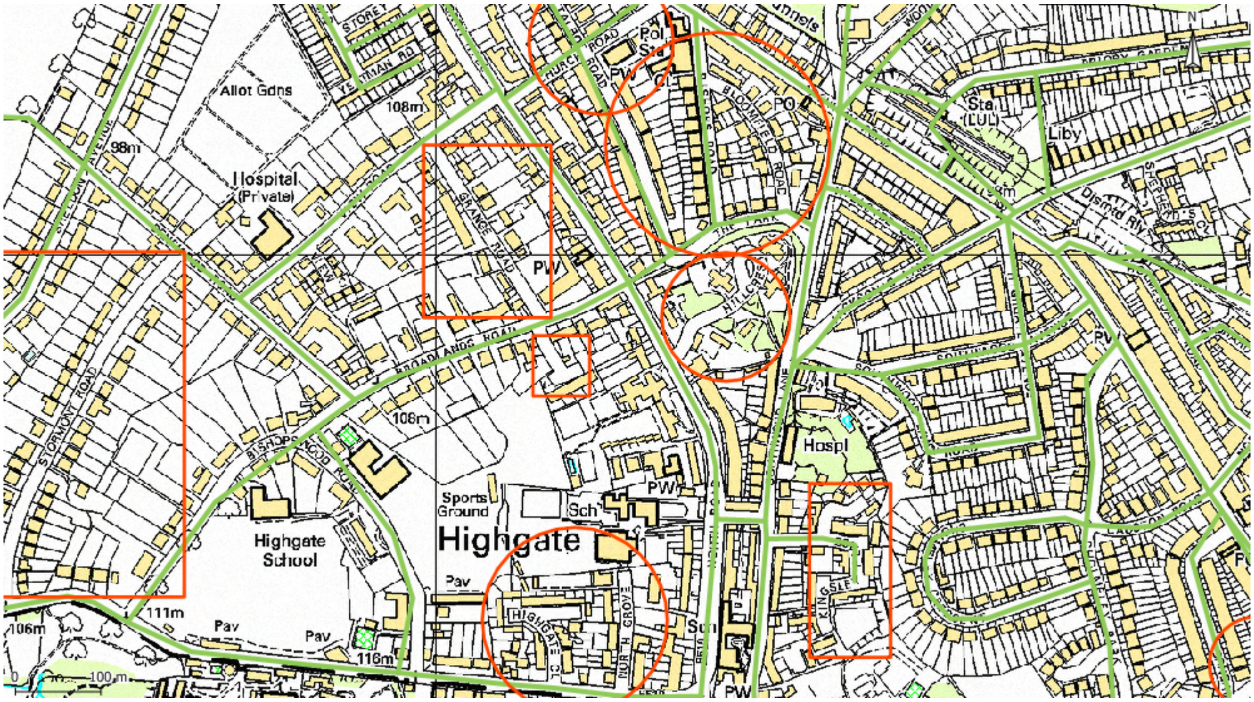


Figure 7 – Highgate area. The light lines are OSM features, and the dark circles and rectangles indicate omissions.

A similar examination of the South Norwood tile, on the other hand, shows small areas that are missing completely, while the quality of the coverage is high.

4.5 Social justice and OSM dataset

Another measure of data collection is the equality in which it is collected. Following the principle of universal service, governmental bodies and organisations like the Royal Mail or the Ordnance Survey are committed to providing full coverage of the country, regardless of the remoteness of the location or the socio-economic status of its inhabitants. As OSM relies on the decisions of contributors about the areas that they would like to collect, it is interesting to evaluate the level in which deprivation influences data collection.

For this purpose, the UK government’s Index of Deprivation 2007 (ID 2007) was used. ID 2007 is calculated from a combination of governmental datasets and provides a score for each Super Output Area (SOA) in England, and it is possible to calculate the percentile position of each SOA. Each percentile point includes about 325 SOAs. Areas that are in the bottom percentiles are the most deprived, while those at the 99th percentile are the most affluent places in the UK.

Following the same methodology that was used for completeness, the road datasets from OSM and from Meridian were clipped to each of the SOAs for the purpose of comparison. In addition, OSM nodes were examined against the SOA layer.

As Figure 8 shows, a clear difference between SOA at the bottom of the scale (to the left) and at the top can be seen. While they are not neglected, the level of coverage is far lower, even when taking into account the variability in SOA size. Of importance is the middle area, where a hump is visible – this is

due to the positioning of most rural SOAs in the middle of the ranking and therefore the total area is larger.

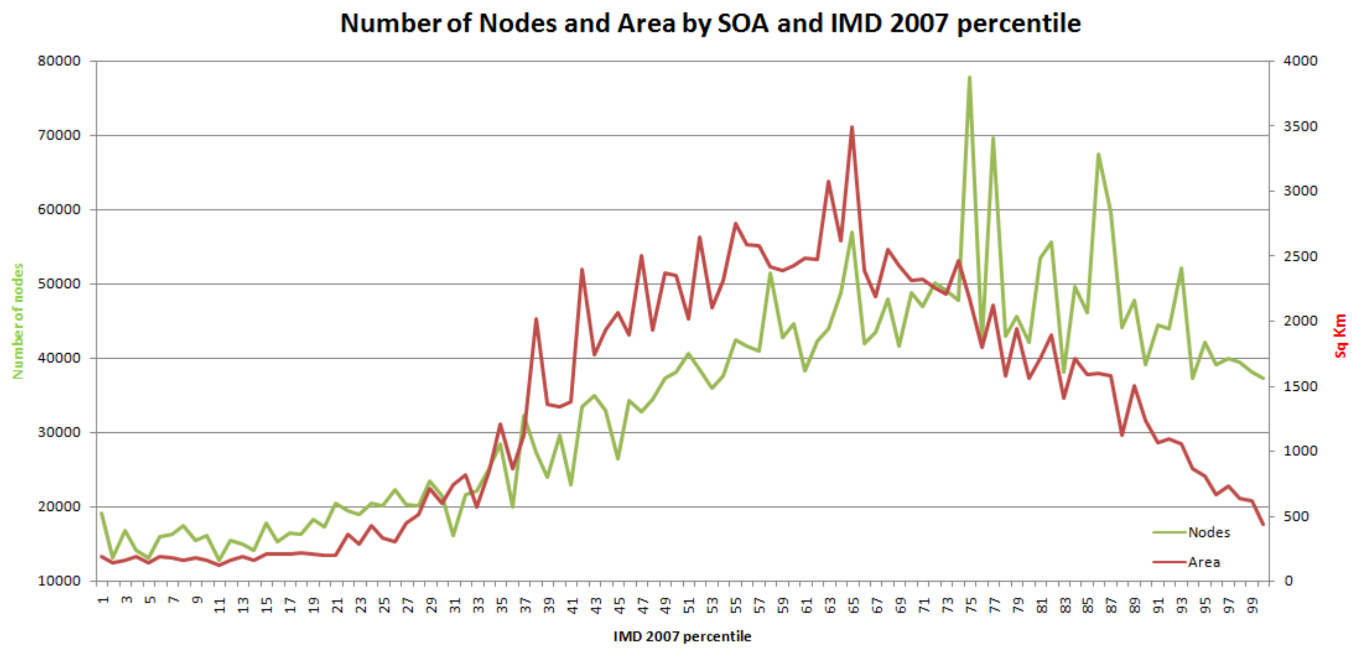


Figure 8 – Number of Nodes and Area by SOA and ID 2007 percentile. Notice that the area of each percentile, about 325 SOAs, is summed up in sq km on the right, while the number of nodes is on the left.

However, nodes provide only partial indication of what is actually captured. A more accurate analysis of mapping activities is to measure the places where OSM features were collected. This can be carried out in two ways. Firstly, all the roads in the OSM dataset can be compared to all the roads in the Meridian dataset. Secondly, a more detailed scrutiny would include only lines with attributes that make them similar to Meridian features, and would check that the name field is also completed – confirming that a contributor physically visited the area as otherwise they would not be able to provide the street name. The reason for this is that only out-of-copyright maps can be used as an alternative source of street names, but they are not widely used as the source of street name by most contributors. Contributors are asked not to copy street names from existing maps, due to copyright issues. Thus, in most cases the recording of a street name is an indication of a physical visit to the location.

These criteria reduce the number of road features included in the comparison to about 40% of the objects in the OSM dataset. Furthermore, to increase the readability of the graph, only SOAs that fall within one standard deviation in area size were included. This removes the hump effect of rural areas, where the SOA size is larger, and therefore allows for a clearer representation of the information.

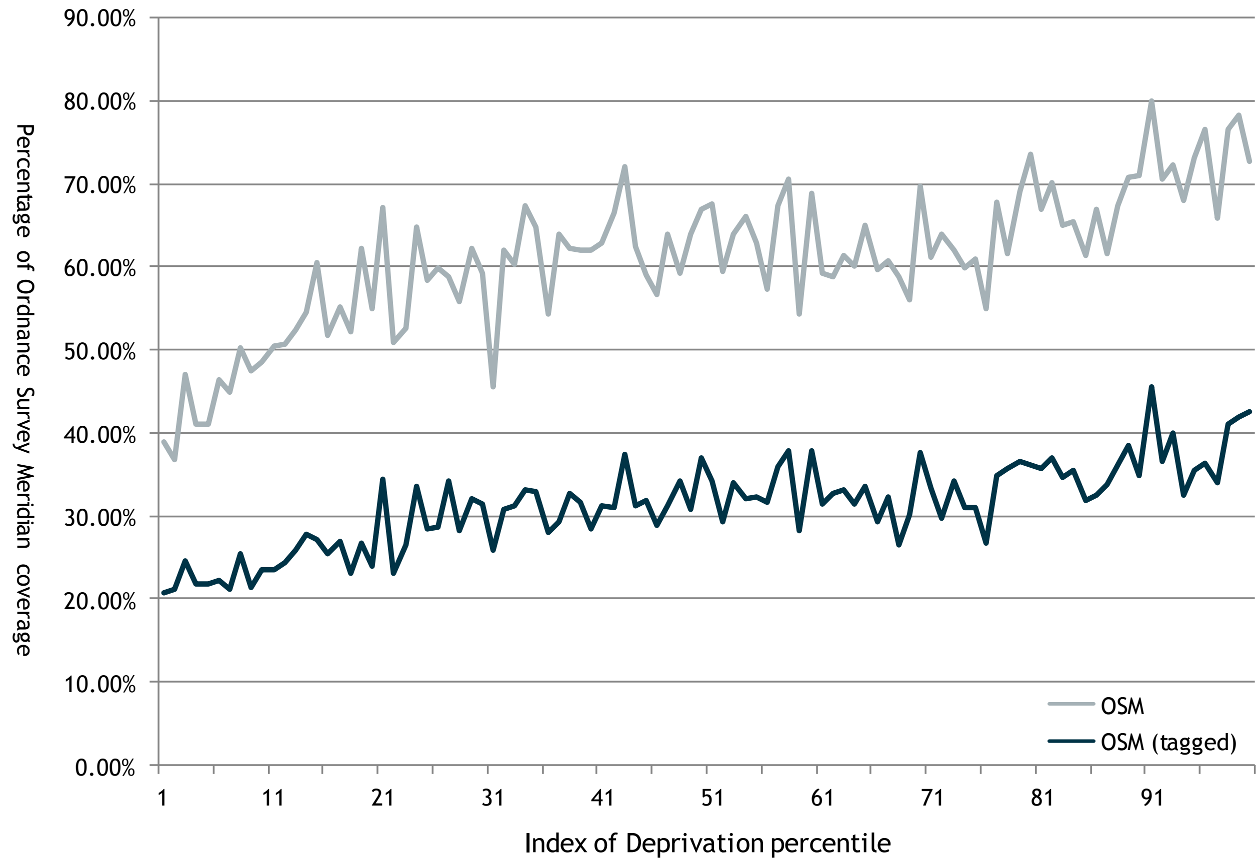


Figure 9 – Percentage of Meridian coverage for OSM (all roads) and OSM (named roads only) by ID 2007 percentile.

Notice that, while the datasets exhibit the same general pattern of distribution as in the case of area and nodes, the bias towards more affluent areas is clearer – especially between places at the top of the scale. As Table 5 demonstrates, at the bottom of the ID 2007 score the coverage is below the average for all SOAs and in named roads there is a difference of 8% between wealthy areas and poor areas.

ID 2007 percentile	All Roads	Named roads
1-10 (poor)	46.09%	22.52%
91-100 (wealthy)	76.59%	30.21%
Overall	57.00%	16.87%

Table 5 – Average percentage coverage by length in comparison to Meridian and ID 2007

This bias is a cause of concern as it shows that OSM is not an inclusive project, shunning socially marginal places (and thus people). While OSM contributors are assisting in disaster relief and humanitarian aid (Maron, 2007), the evidence from the dataset is that the concept of 'Charity begins at home' has not been adopted yet. This indeed verifies Steve Coast's declaration that 'Nobody wants to do council estates. But apart from those socio-economic barriers – for places people aren't that interested in visiting anyway – nowhere else gets missed' (GISPro, 2007).

Significantly, it is civic society bodies such as charities and voluntary organisations that are currently excluded from the use of the commercial dataset due to costs. The areas at the bottom of the Index of Deprivation are those that are most in need in terms of assistance from these bodies, and thus OSM is failing to provide a free alternative to commercial products where it is needed most.

5. *Spatial data quality and VGI*

The analysis that was carried out here exposed many aspects of the OSM dataset, which can provide an insight to the quality of VGI in general. The most impressive aspect is the speed in which the dataset was collected – within a short period of time, about a third of the area of England was covered by a team of about 150 participants with minor help from over 1000 others. The availability of detailed imagery is crucial, as can be seen from the impact of high-resolution Yahoo! imagery (Figure 4). The matrix of positional accuracy shows that OSM information is of a reasonable accuracy of about 6 metres, and a good overlap of up to 100% of Ordnance Survey digitised motorways, A-roads and B-roads. In places where the participant was diligent and committed, the information quality can be, indeed, very good.

This is an interesting aspect of VGI and demonstrates the importance of the infrastructure, which is funded by the private and public sector and which allows the volunteers to do their work without significant personal investment. The GPS system and the receivers allow untrained users to automatically acquire their position accurately, and thus simplify the process of gathering geographical information. This is, in a way, the culmination of the process in which highly trained surveyors were replaced by technicians, with the introduction of high-accuracy GPS receivers in the construction and mapping industries over the last decade. The imagery also provides such an infrastructure function – the images were processed, rectified and georeferenced by experts and thus, an OSM volunteer who uses this imagery for digitising benefits from the good positional accuracy which is inherent in the image. So the issue here is not to compare the work of professionals and amateurs, but to understand that the amateurs are actually supported by the codified professionalised infrastructure and develop their skills through engagement with the project.

At the same time, the analysis highlights the inconsistency of VGI in terms of its quality. Differences in digitisation – from fairly sloppy in the area of Highgate to consistent and careful in South Norwood – seem to be part of the price that is paid for having a loosely organised group of participants. Figure 2 also shows a range of performances when compared to A- and B-roads, and there is no direct correlation between the length of the road and the quality of its capture.

This bring to the fore the statement that Steve Coast has made, regarding the need to forgo the concept of completeness. Indeed, it is well known that because of update cycles, cartographic limitations such as projections and a range of other issues are leading to uncertainty in centrally collected datasets such as those created by the Ordnance Survey of the United States Geological Survey (Goodchild, 2008). Notice, for example, that, despite the fact that Meridian is generalised and does not include minor roads, this does not diminish its usability for many GIS analysis applications. Moreover, with OSM, in terms of dealing with incompleteness, if users finds that data is missing or erroneous, they do not need to follow a lengthy process of error reporting to the data producer and wait until it provides a new dataset; rather they can fix the dataset themselves and, within a short period of some hours, use the updated and more complete dataset (see OSM 2008).

Yet, there are clear differences between places that are more complete and areas that are relatively empty. A research question that is emerging here is about the usability of the information – at which point does the information become useful for cartographic output and general GIS analysis? Is there a

point over which the coverage is good enough? Or is coverage of main roads in a similar fashion to Meridian enough? If so, then OSM is more complete than was stated above – likely to be at near 50%. These are questions that were partially explored during the 1980s and 1990s in the spatial data quality literature, but a more exploratory investigation of these issues for VGI is necessary.

Another core question that the comparison raises, and which Goodchild's (2008) discussion is hinting at, is the difference between declared standards of quality, and ad hoc attempts to achieve quality. In commercial or government-sponsored products, there are defined standards for positional accuracy, attribute accuracy, completeness and other elements that van Oort (2006) listed. Importantly, the standard does not mean that the quality was achieved for every single object – for example, a declaration of positional accuracy of a national dataset that it is within 10m on average from its true location means that some objects might be as far as 20m from their true location. All we have is a guarantee that, overall, map objects will be within the given range and this is based on trust in the provider and its quality assurance procedures.

In the case of VGI, there is clear awareness of the quality of the information. For example, Cherdlu (2007) presentation on quality in the first OSM community conference; or the introduction in June 2008 of OpenStreetBugs – a simple tool that allows rapid annotation of errors and problems in the OSM dataset; and the development by Dair Grant, an OSM contributor, of a tool to compare Google Maps and OSM to identify possible gaps – similar to the process that was described in Section 4.4 (see http://www.refnum.com/osm/gmaps/). However, in the case of OSM, unlike Wikipedia or Slashdot (a popular website for sharing technology news), there is no integrated quality assurance mechanism that allows participants to rate the quality of the contribution of other participants. This means that statements about accuracy, such as the one discussed here, come with a caveat. Looking at Figure 2, the following statement can be formulated: 'you can expect OSM data to be with positional accuracy of over 70%, with occasional drop down to 20%'. In terms of overall quality, this might lead to results that are not dissimilar to commercial datasets, apart from a very significant difference: while our expectation from the commercial dataset is that errors will be randomly distributed geographically, sections 4.2 and 4.4 highlighted the importance of the specific contributor to the overall quality of the information captured in a given area. Therefore, the errors are not randomly distributed. This raises a question about the ways in which it is possible to associate individual contributors with some indication of the quality of their outputs. Another interesting avenue for exploration is emerging from the principle of Open Source software development, which highlights the importance of 'Given enough eyeballs, all bugs are shallow' (Raymond, 2001, p.19). For mapping, this can be translated as the number of contributors that worked on an area and therefore removed 'bugs' from it. Is this indeed the case? Are areas that are covered by multiple contributors exhibiting higher positional and attribute quality?

The analysis also highlighted the implications of the digital and social divide on VGI. Notice the lack of coverage in rural areas and poorer areas. Thus, while Goodchild (2007b) suggested that 'the most important value of VGI may lie in what it can tell about local activities in various geographic locations that go unnoticed by the world's media, and about life at a local level', the evidence is that places that are perceived as 'nice places', where members of the middle classes have the necessary educational attainment, disposable income for equipment and availability of leisure time, will be covered. Places where population is scarce or deprived are, potentially, further marginalised by VGI exactly because of the cacophony that the places which are covered create.

There are many open questions that this preliminary investigation could not cover. First, within the area of spatial data quality there are several other measures of quality that were mentioned in section 2, and that are worth exploring with VGI – these include logical accuracy, attribute accuracy, semantic accuracy and temporal accuracy. The temporal issue is of special interest with VGI; due to the leisure-activity

aspect of the involvement in such projects, the longevity of engagement can be an issue as some participants can get excited about a project, collect information during the period when the map is empty, and then lose interest. OpenStreetMap is still going through a period of rapid growth and the map is relatively empty, so this problem has not arisen yet. However, the dataset allows the exploration of this issue and there is a need to explore the longevity of engagement. It is important to note that many other commons-based peer-production projects are showing the ability to continue and engage participants over a long period – as shown by the Apache web server, which has been developing for almost 15 years, or Wikipedia, which continues to engage its participants after eight years.

This preliminary study has shown that VGI can reach very good spatial data quality. As expected in this new area of research, the analysis opens up many new questions about quality – from the need to explore the usability of the data, to the consistency of coverage and the various elements of spatial data quality. Because the activity is carried out over the Internet, and in projects like OpenStreetMap the whole process is open to scrutiny at the object level, it can offer a very fruitful ground for further research. The outcome of such an investigation will be relevant beyond VGI, as it can reveal some principles and challenge some well-established assumptions within GIScience in general.

Acknowledgements

Many thanks to Patrick Weber, Claire Ellul, and especially Naureen Zulfiqar and Aamer Ather who carried out part of the analysis of motorways as part of their M.Eng. project. Some details on geographical information quality are based on a report on the Agency Performance Monitor that was prepared for the Ordnance Survey in 2005. Some early support for OSM research was received from the RGS Small Research Grants programme in 2005. Apologies to the OSM system team for overloading the API server during this research.

The author is grateful to Steven Feldman, Steve Coast, Mike Goodchild, Martin Dodge and four anonymous reviewers who provided comments on an earlier version of this paper.

Thanks to the Ordnance Survey External Research and University Liaison who provided the Meridian 2 dataset for this study in January 2008. The 1:10,000 raster and the SOA boundaries were provided under EDINA/Digimap and EDINA/UKBorders agreements. ID 2007 was downloaded from DCLG and is Crown copyright.

References

Benkler, Y. and Nissenbaum, H. (2006) Commons-Based Peer Production and Virtue. *Journal of Political Philosophy*, 14 (4): 394-419.

Burrough, P.A. and Frank, A.U. (eds) (1996) Geographic objects with indeterminate boundaries. London: Taylor & Francis, p.345.

Cherdlu, E. (2007) *OSM and the art of bicycle maintenance*, Presented at State of the Map 2007, 14-15 July, Manchester, UK, http://www.slideshare.net/nickb/openstreetmap-and-the-art-of-bicycle-maintenance-state-of-the-map-2007/

Chrisman, N.R. (1984) The role of quality information in the long-term functioning of a geographic information system. *Cartographica*, 21(2-3), 79-87.

Chrisman, N.R. (1991) The error component in spatial data, Chapter 12, p. 165-174 in Maguire, D.J., Goodchild M.F. and Rhind, D.W. (eds) *Geographical Information Systems: Overview Principles and Applications*, Longmans.

Cross, P., Haklay, M. and Slingsby, A. (2005) Agency Performance Monitor Audit – Consultancy Report, UCL, London.

Elwood, S. (2009) Geographic Information Science: new geovisualization technologies – emerging questions and linkages with GIScience research, *Progress in Human Geography* 33(2) pp. 256–263

Fisher, P.F. (1999) Models of uncertainty in spatial data. In: Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (eds), *Geographical Information Systems: Principles and technical issues*. New York: John Wiley & Sons, 2nd ed., vol. 1, pp. 191-205.

Friedman, Thomas L. (2006) *The world is flat: A brief history of the twenty-first century, updated and expanded edition*. New York: Farrar, Straus and Giroux.

GISPro (2007) *The GISPro Interview with OSM founder Steve Coast*, GIS professional, Issue 18, October 2007, pp. 20-23.

Goodchild, M.F. (2008) Assertion and authority: the science of user-generated geographic content. Proceedings of the Colloquium for Andrew U. Frank's 60th Birthday. GeoInfo 39. Department of Geoinformation and Cartography, Vienna University of Technology.

Goodchild, M.F. (2007a) Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0, *International Journal of Spatial Data Infrastructures Research*, Vol. 2, 24-32.

Goodchild, M.F. (2007b) Citizens as sensors: the world of volunteered geography. *GeoJournal*, Vol. 69, No. 4, pp. 211-221.

Goodchild, M. F. and Hunter, G. J. (1997) A simple positional accuracy measure for linear features, *International Journal of Geographical Information Science*, 11(3): 299-306.

Goodchild, M. F. (1995) Attribute accuracy. In: Guptill, S.C. and Morrison, J.L. (eds), Elements of Spatial Data Quality. Amsterdam: Elsevier, pp. 59–79.

Goodchild, M.F. (1992) Geographic Information Science, *International Journal of Geographical Information Systems*, 6(1), 31-45.

Goodchild, M.F., Haining, R. and Wise, S. (1992) Integrating GIS and spatial data analysis: problems and possibilities, *International Journal of Geographical Information Science*, 6(5), 407-423.

Haklay, M. and Weber, P. (2008) OpenStreetMap – User Generated Street Map, *IEEE Pervasive Computing.*

Haklay, M., Singleton, A. and Parker, C. (2008) Web mapping 2.0: the Neogeography of the Geoweb, *Geography Compass.*

Howe, J. (2006) The Rise of Crowdsourcing, *Wired Magazine,* June 2006.

Hunter, G.J. (1999) New tools for handling spatial data quality: moving from academic concepts to practical reality, *URISA Journal*, 11(2), 25-34.

Keen, A. (2007) *The Cult of the Amateur*, Broadway Business.

Kresse, W. and Fadaie, K. (2004) *ISO Standards for Geographic Information*, Springer.

Maron, M. (2007) *OpenStreetMap: A Disaster Waiting to Happen*, Presented at State of the Map 2007, 14-15 July, Manchester, UK, http://www.slideshare.net/mikel_maron/openstreetmap-a-disaster-waiting-to-happen

Nielsen, J. (2006) Participation Inequality: Encouraging More Users to Contribute, Alertbox, 9th October 2006.

Oort, P.A.J. van (2006) Spatial data quality: from description to application, PhD Thesis, Wageningen: Wageningen Universiteit, p. 125.

Openshaw, S. (1989). Learning to live with error in spatial databases. In Goodchild, M. F. and Gopal, S. (eds), Accuracy of Spatial Databases. London: Taylor & Francis.

Ordnance Survey (2007) Meridian 2 User Guide and Technical Specification, Version 5.2 February 2007, Ordnance Survey, Southampton

OSM (2008) *Places*, Available http://wiki.openstreetmap.org/index.php/Places [Accessed 28th July 2008]

Perkins, C. and Dodge, M. (2008) The potential of user-generated cartography: a case study of the OpenStreetMap project and Mapchester mapping party, *North West Geography*, 8(1), 19-32.

Raymond, E.S. (2001) The Cathedral and the Baazar, O'Reilly.

Sui, D. Z. (2008) The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. *Computers, Environment and Urban Systems* 32, pp. 1–5.

Tapscott, D. and Williams, A.D. (2006) Wikinomics – How mass collaboration changes everything. London: Atlantic Books.

Where 2.0 (2008) Where 2.0 Homepage http://en.oreilly.com/where2008/public/content/home [Accessed 23rd July 2008]