The Annotation Process in OpenStreetMap[Not for distribution]

Peter Mooney[1] and Padraig Corcoran[2]

1. Department of Computer Science, National University of Ireland Maynooth,
Maynooth,Co. Kildare. Ireland.
e-mail: {peter.mooney}@nuim.ie
2. School of Computer Science and Informatics, University College Dublin,
Belfield, Dublin 4. Ireland.
e-mail: {padraig.corcoran}@ucd.ie

Abstract: In this paper we describe the analysis of $25,000$ objects from the OpenStreetMap (OSM) databases of Ireland, United Kingdom, Germany, and Austria. The objects are selected as exhibiting the characteristics of "heavily edited" objects. We consider "heavily edited" objects as having 15 or more versions over the object's lifetime. Our results indicate that there are some serious issues arising from the way contributors tag or annotate objects in OSM. Values assigned to the "name" and "highway" attributes are often subject to frequent and unexpected change. However, this "tag flip-flopping" is not found to be strongly correlated with increasing numbers of contributors. We also show problems with usage of the OSM ontology/controlled vocabularly. The majority of errors occurring were caused by contributors choosing values from the ontology "by hand" and spelling these values incorrectly. These issues could have a potentially detrimental effect on the *quality* of OSM data while at the same time damaging the perception of OSM in the GIS community. The current state of tagging and annotation in OSM is not perfect. We feel that the problems identified are a combination of the flexibility of the tagging process in OSM and the lack of a strict mechanism for checking adherence to the OSM ontology for specific core attributes. More studies related to comparing the names of features in OSM to recognised ground-truth datasets are required.

1 Introduction

OpenStreetMap (OSM) is a collaborative project to create a free editable map of the world. It is currently probably the most prominent and well-known example of Volunteered Geographic Information (VGI) (Goodchild, 2007, 2008) on the Internet. The OSM database is a very significant collection of volunteer collected spatial data and is worthy of research investigation. OSM is based on the "wiki collaborative model". Prasarnphanich and Wagner (2011) remarks that the wiki model's readwrite web paradigm, which enables peer production and incremental improvement in an integral and organic way "has led to the creation of significant knowledge assets and corresponding knowledge communities". The OSM Statistics page (OSM-Stats, 2011) on the OSM wiki shows, in real-time, the number of users, GPS traces, nodes, ways, and relations currently stored in the OSM database. At the time of writing (July 2011) there are over 100 million ways in the database. There are over

$400,000$ contributors registered in the OSM project. Volunteers in the OSM community collect geographic information using GPS devices and submit this to the global OSM database (Ciepluch et al., 2009). In recent years companies such as Yahoo! and Bing have made global aerial imagery available to the OSM projects. This imagery can be used as a base layer in one of several OSM editors where volunteers can trace the outline of geographical features from the aerial imagery. Spatial datasets which are available under OSM compatible free and open data licenses can be "bulk" imported into OSM. Examples include: Automotive Navigation Data donation of the entire road and street network database of the Netherlands, CORINE Landcover databases for France and Estonia, and the TIGER road network dataset in the United States. However, for many reasons, OSM discourages bulk importing unless it is supported by the larger OSM community (see OpenStreetMap (2011*b*)). For example if bulk imports are incorrectly managed they may delete existing contributed data within the OSM database in areas where specific OSM contributors are actively maintaining this data. The preference and priority for spatial data collection, importing, and editing of data in the OSM global database is with 'volunteer mappers' (individual and groups).

Real world geographic features are represented in OSM as points, lines, and polygons. Thematic attributes for these features are stored as *tags*. Tagging has emerged as a popular means to annotate online objects such as bookmarks, photos, and videos (Cantador et al., 2011). In most collaborative systems, users create or upload content (items), annotate them with freely chosen words (tags), and share it with other users (who may in turn edit or update the annotations). In OSM there are no upper limits to the number of tags associated with any object. While it is discouraged, and editing software will identify the problem, objects do not necessarily have to be assigned any tags. OSM does not have any content restrictions on tags that can be assigned to points, lines, or polygons. One can use any tags provided "the values are verifiable" (OpenStreetMap, 2011*a*). The Map Features guide (OpenStreetMap, 2011*a*) emphasises that there is "a benefit in agreeing to a recommended set of features and corresponding tags in order to create, interpret and display a common basemap". Using tags (and their recommended set of values) increases the likelihood that spatial data contributed to OSM will be understood by various cartographic rendering engines which create map visualisations from OSM data. OSM contributors can also tag and edit objects that they did not create themselves. Tagging has been described as "one of the dilemmas in user behaviour in Web 2.0" (Liu et al., 2011). Properly and exhaustively tagging all objects (for example objects in OSM or photographs in an online album) is labour intensive and time-consuming but is important for the overall quality of the collection of objects. Poor or incorrect tagging leads to unsatisfactory results (Liu et al., 2011). In this paper we use "annotation" as a synonym for tagging or assigning tags to objects in OSM.

This paper gives an overview of characteristics of the annotation process in OSM to investigate what happens when multiple contributors edit objects in a spatial database. The paper shows that OSM data is constantly changing and evolving and through analysis of the history of the evolution of objects one can see how objects have changed (in response to edits) over time. OSM is a very large (and growing) spatial dataset and probably the most well known example of VGI. We analyse the annotation of $25,000$ spatial objects or features in the OSM databases of Ireland (653), United Kingdom ($10,040$), Germany ($10,604$), and Austria ($3,367$). Extracting the history of objects in OSM is currently a difficult and time consuming process (Mooney and Corcoran, 2011*b*,*a*). We specifically selected objects which are "heavily edited": that is objects with 15 or more versions of edits. We set V as the lower threshold of versions at $V \geq 15$. These objects are more likely to exhibit collaborative editing where multiple OSM contributors edit and annotate these objects. We feel these objects are particular interesting and could eventually assist us in understanding the nature of contributions in the OSM collaborative project model. Antin (2011) remarks that for collaborative projects (in their case Wikipedia) focus often quickly turns to the “practical challenges of information quality, coordination, and contributor bias” related to these open models.

The remainder of this paper is organised as follows. Section 2 provides a discussion of the literature related to the topic of contributors to VGI projects and the quality of their contributions. In Section 3 we discuss the experimental setup for this research. It is necessary to describe how the $25,000$ objects were selected from OSM and how their histories were compiled. The core of the paper is Section 4 which outlines the results of our analysis on the selected $25,000$ OSM objects. Section 5 summarises the key outcomes from the paper and in this section we present some issues for future work and research on this topic.

2 Overview of Related Work

In this section we provide an overview of related work. To better organise this overview we have divided related work into literature-based related work (section 2.1) and then web-based related work (section 2.2). Web-based related work is for research which may not have peer-reviewed literature related to it.

2.1 OSM in peer-reviewed literature

VGI and OpenStreetMap are exciting research areas at present (Mooney et al., 2010*b*) in GIS and related disciplines (Goodchild, 2008). Qian et al. (2009) remarks that “since general users can add and change data in VGI, the stored data should update frequently, and result in an abundant and updated geographic dataset”. This has “reversed the traditional top-down flow of information” and Flanagin and Metzger (2008) state that as the amount of VGI continues to grow “the issues of credibility and quality should assume a prominent place on the research agenda”. Without some quantitative measures of accessing the quality of the OSM data the GIS community has been slow to consider OSM as a serious source of data (Mooney et al., 2010*b*). While spatio-temporal accuracy and quality are fundamental requirements for GIS modelling and applications, documentation, metadata, and attribution of data is also of major importance. This problem is experienced in almost all domains. Bulterman (2004) suggests that the “complete disregard for documentation of data resources” has made it almost impossible for one to perform a fitness for use or fitness for purpose evaluation on data resources on the Internet. Brando and Bucher (2010) advise that the quality of VGI is enhanced if proper metadata is created and maintained containing information on: types of changes and edits, methods of survey and collection, and finally a fitness for purpose statement.

Many papers in the literature report very positive experiences and results for OSM. Haklay (2010) describes a comparison of the road network in OSM for England with the road network in the Ordnance Survey UK Meridian dataset. The conclusions of Haklay’s study is that OSM is “as good if not better than the Meridian dataset in terms of positional accuracy”. Haklay remarks that “completeness is very good for major urban centres and draws the conclusion that if mapping applications want to use OSM data for these locations it is as good a choice as any other source of spatial data. A similar study by Zielstra and Zipf (2010*a*,*b*) of OSM and TeleAtlas for Germany shows that for larger cities (Berlin, Frankfurt, Munich, etc) the OSM spatial data “is so rich that OSM is now replacing proprietary data for many projects”. Ludwig et al. (2011) compare Navtec and OSM street and road networks also for Germany and conclude that between the two datasets there “are considerable qualitative differences between regions, towns, and street categories" but at a national level the "relative completeness of OSM objects is high enough for maps and cartographic production". Zielstra and Hochmair (2011) compare OSM, TIGER, NAVTEQ NAVSTREETS and Tele Atlas Multinet street data for the state of Florida, USA in a related study to Zielstra and Zipf (2010*a*,*b*). Zielstra and Hochmair (2011) find "OSM strong heterogeneity of OpenStreetMap data for the US in terms of its completeness". Ather (2009) comments that as OSM grows, most regions (the UK in their study) will eventually fulfil the levels of map quality required for other GIS applications. He goes on to comment that "it would be useful if long term measures were in place to provide continued assessments of OSM map quality and then communicate these results back to users as they browse through the map". In Over et al. (2010) the authors comment that the quality control of OSM differs fundamentally from professionally edited maps. The community-based approach allows anyone to upload and alter the spatial data. But due to the huge number of editors, errors and conflicts are usually quickly resolved. They state that "OSM has probably the most up-to-date map data". In urban areas, changes in the road network appear in the OSM data set long before appearing in other map data providers. Haklay et al. (2010) investigates the relationship between the number of contributors to OSM and data accuracy against a ground-truth dataset. Haklay et al. conclude that, beyond 15 contributors per square kilometre, the positional accuracy of OSM becomes very good (below 6 metres). At the other end of the scale, the first 5 contributors to an area seem to provide the biggest contribution in terms of positional accuracy improvement. Girres and Touya (2010), in their quality assessment of OSM dataset for France, show that the number of OSM objects in an area clearly grows in relation to the number of contributors in the area but under a non-linear relationship. As is clearly shown here OSM is a multiple representation database containing Points-of-Interest, land cover, transportation networks, buildings, waterways and waterbodies. There is also some literature on the nature of contributors to VGI and OSM. Coleman et al. (2009) show the participants in the production process of VGI both users and producers or "prosumers". Assessing the credibility of contributors is important to evaluating the overall reliability of their contribution. They find many reasons why contributors take part in VGI including: for social rewards, to take part in an outlet for creative and independent self-expression, pride in one's home place, and intellectual stimulation. Budhathoki et al. (2008) argue that motivations of the contributors from VGI are very strong and can assist in the "distribution of the production of GI for Spatial Data Infrastructures (SDI) among organizations, individuals, and groups of individuals". Then a "hybrid SDI model that draws on the synergy between the conceptual foundation of SDI and an extensive produser base of VGI can be developed".

However, there are also several negative outcomes for OSM from some of these research studies. Welser et al. (2011) remarks that there is the perception amongst many in the scientific domain that the quality of open collaborative projects, such as Wikipedia, "will never been sufficient as long as it relies on non-expert volunteers of unknown identity" and this appears to be an issue for some in the GIS community regarding OSM. Qian et al. (2009) conclude that a serious negative aspect of the VGI model is that the underlying data is acquired by non-professionals with non-professional equipment meaning that there cannot be any guarantee of quality about the VGI (or OSM) data unless it can be compared to some other source. Ballatore and Bertolotto (2011) call OSM "spatially-rich but semantically-poor". Smart et al. (2011) show how freely available sources of georeferenced data can be used for automated enrichment of 3D city models. OSM is included as a key data source. They concluded that "matching the georeferenced point locations from sources, such as OSM, to the geometry of the buildings in the registered 3D model" posed significant problems due to accuracy and sparse attribute problems. While Haklay (2010)'s comparison of OSM and Ordnance Survey UK data reflects very positively on OSM the author concludes by warning that "there are serious issues about completeness and coverage" in the UK. Coverage is also commented upon in the work of Zielstra and Zipf (2010*a*,*b*) who state that "while professional data is not without it's faults the coverage of OSM in rural areas is too small to be seriously considered a sophisticated alternative for *any* applications". When one moves away from large urban centres the major issue for quality becomes one of coverage - in many rural areas there is little or no OSM coverage at all. While Ludwig et al. (2011)'s comparison of Navtec and OSM street/roads is positive from the OSM perspective the authors conclude that "other attributes of OSM, which are needed for other advanced GIS problems, are still relatively incomplete". Mooney and Corcoran (2011*c*) investigate the potential role VGI can play in eEnvironment and various Spatial Data Infrastructures (SDI) on a local, regional, and national level. Specifically for OSM the authors conclude that while currently problems such as inconsistency of metadata and unpredictable changes to geometries are a barrier to inclusion in SDI the quantity of spatial data in OSM means it has a role to play in SDI. Mooney et al. (2010*a*) investigate the spatial representation of natural features in OSM. They report that there are differences in polygon structure for natural and landuse polygons based on: sampling point density, simplification and generalisation of imported data, and inconsistency in manual tracing from aerial imagery. Overall this highlights inconsistencies in representation of natural features in OSM databases for different countries, regions, and contributors. Mooney et al. (2010*b*) apply shape matching techniques from pattern recognition to compare OSM polygons (lakes) with Ordnance Survey Ireland NMA data. Their results reveal that the shapes of these polygons in OSM compare poorly to authoritative NMA data. The authors conclude that the quality of OSM data is not necessarily solely restricted to a geometric comparison to some other dataset but should include other aspects such as metadata and tagging. Girres and Touya (2010) also suggest that tagging and annotation of objects within OSM deserves immediate attention.

The production of cartographic output from the OSM database is the most popular use of the raw spatial data with some authors (Kessler et al., 2009; Over et al., 2010) remarking that OSM is not considered for "serious Geomatics applications". GIS-based research using OSM as input data for models and testing is beginning to appear in the literature. Wallgrün et al. (2010) use OSM for matching a sketch map to a geo-referenced data set. Corcoran et al. (2011) use OSM data for testing map simplification models. The same authors use OSM data for model testing in their work on progressive transmission of vector data (Corcoran and Mooney, 2011). Jacob et al. (2010) use OSM as the source of spatial data for routing algorithms used in the development of haptic-feedback applications for mobile pedestrian navigation systems.

2.2 OSM-driven Web-based applications and research

Some related work which uses the history of objects in OSM is available as web services online. While *MapCompare* (Geofabrik, 2011*b*) is not a history browsing tool, it allows visual comparison of OSM with Google Maps, Bing, etc. Snapshot images could be compiled over time to provide a visual "history" of a specific area or feature(s). The *OSM History Browser* by Langläufer (2011) provides a simple interface to retrieve the entire version history of any OSM object (node, way, or relation). The OSM identification number of the object is required. The version history of the object is provided in HTML table format. Differences between versions can be inspected using the "compare versions" function. The *OSM History Viewer* by OSM History Viewer (2011) is a similar web application. For this web application a little more knowledge about the internal data management of OSM is required as one must supply the identification number of "changesets" to the application. A changeset is a collection of all the edits performed by a particular user over a 24 hour period. The *OSM History Coverage viewer* by

Ramm (2011) is a web-based service that creates animated GIF images that depict how OSM coverage of an area has changed over time. Computation is too time-consuming to offer this as a live service, but one can request images to be created and then view or download them once they are ready. Trame and Kessler (2011) describes a web-application which generates heat-maps for nodes in OSM. The version number of the nodes (and polygons) is used in heat-map visualisation. Roick (2011) creates visualisations of OSM data for Europe in hexagonal cells. Version numbers is one of the attributes visualised. van Exel et al. (2010) consider if version history could be used as a variable in building trust and quality metrics for OSM data. Similar work is appearing for Wikipedia but this work is at a more advanced and mature stage of development. *iChase*, developed by Riche et al. (2010), visualizes the trend of activities for articles and contributors. It allows users to interactively explore the history of changes by drilling down into specific articles and contributors, or time points to access the details of the changes. In similar work Suh et al. (2008) describe WikiDashBoard which also provides drill down functionality on the history of Wikipedia articles Pirolli et al. (2009) claims that their user trails using WikiDashBoard suggested that "increased exposure to the editing/authoring histories of Wikipedia increases credibility judgements by users". In the next section we describe the experimental design and setup and discuss the characteristics of spatial objects in OSM.

3 Experimental Setup

In this section we provide an overview of the experimental setup for the analysis presented in Section 4. In section 3.1 we discuss how the OSM data is obtained, processed, and prepared for analysis. In section 3.2 we discuss the process of selecting OSM objects for our analysis. Section 3.3 discusses the characteristics of the selected objects and regions from which they are drawn. As the OSM global database contains several million objects we feel that it is necessary to carefully select a subset of these objects for analysis.

3.1 Understanding the OSM data

OSM data is freely available, in OSM-XML format and ESRI Shapefile format, from the GeoFabrik web service (see Geofabrik (2011*a*)). This data is updated almost hourly. Consequently, the most up-to-date version of the OSM database for any region of the world is always available. GeoFabrik provide the OSM data divided into country and continent *packages* which makes it very easy to download specific regions of interest rather than processing the enormous and rather unwieldy entire "planet.osm" dataset for the global OSM. The "planet.osm" history dump file (OSM-XML format) is available for download. The uncompressed version of this file is currently close to $500Gb$ in size. The enormity of this OSM-XML file makes it difficult to work with both conventional XML processing software packages and programming languages. Hardware issues of disk space availability and memory usage mean that processing this file is beyond the capabilities of most standard desktop or server computers. The OSM API (OSM API, 2011) allows access to the history of nodes, ways, and relations. These are also returned in OSM-XML format. One must make a separate API call for each unique node, way, or relation required. In the case of ways the OSM-XML returned containing the history is structured as follows. Each version of the way is included in chronological order of when it was created. For each version of the way an unordered listing of the tags (key-value pairs) associated with that version is also included. The timestamp and user id of the contributor is included. Unfortunately only the OSM identifiers of the nodes in each way are provided. A separate API call must be performed to lookup and retrieve the spatial coordinates of nodes in each way. This makes the process very time consuming rather than computationally complex. In Mooney and Corcoran (2011*a*) we describe a software-based method for downloading the history of a chosen set of objects from OSM. This involves firstly identifying objects in the most current version of the OSM database. Then for each of these objects their history (in OSM-XML format) is downloaded directly from the OSM servers using the OSM API (OSM API, 2011). The history of each object is only a reference to the nodes used to create each geometry. Subsequently, each node must be downloaded from the OSM API (OSM API, 2011) to create the geometry of the objects in a PostGIS database.

3.2 Selecting OSM Objects for Analysis

The OSM global database contains several million objects (OSM-Stats, 2011). Consequently we felt that it was necessary to carefully select a subset of these objects for analysis. As our emphasis is on tagging and annotation it was necessary to select objects which had a non-empty set of tags and had tags for the most frequently occurring attributes including: name, highway, and landuse. “Heavily edited” objects in OSM should provide good examples of “significant editing and revision work by many contributors” (Anderka et al., 2011). This criteria allows us to discard analysis, for this study, of objects in OSM with a very low number of edits. A closely related concept in Wikipedia to heavily edited objects in OSM is the “featured article” and much research work related to the quality and trustworthiness of Wikipedia articles focus on “featured articles”. In Javanmardi and Lopes (2010) the authors discuss the development of a model for the evolution of content quality in Wikipedia articles in order to estimate the fraction of time during which articles retain high-quality status. They select only “featured articles”. As outlined by Anderka et al. (2011) featured articles in Wikipedia are “made” after significant editing and revision work by many contributors and moderators. Stein and Hess (2007) argue that “instead of just looking at the formal characteristics of featured articles one should look at what contributors do on these pages” in order to understand the effects of multiple contributors over an extended period of time.

3.3 Heavily Edited Objects

Table 1 provides a summary of the distribution of version numbers in the four OSM datasets we have used in the analysis in this paper. The Ireland OSM dataset is used. This includes the Republic of Ireland and Northern Ireland as part of the island of Ireland. The United Kingdom consists of England, Scotland, and Wales. Germany and Austria were chosen for inclusion in this study because they have two of the most active OSM communities in Europe. It is very interesting to note from Table 1 that a very large percentage of ways in all four datasets have 5 or less versions. The percentages are as follows: Ireland (95.3% as $232,707$ from $244,192$), United Kingdom (95.4% as $3,384,643$ from $3,549,831$), Germany (93.1% as $10,445,536$ from $11,226,308$), and Austria (89.0% as $1,085,003$ from $1,219,045$). It was necessary to choose a threshold V on or above which features could be considered as “heavily edited” or “popular” features. Unfortunately, to our current knowledge, there is no similar work in the literature from which we can base this choice. For the purposes of this work we choose to set the threshold value V as 15. Setting V as 15 should allow us to gather features from the OSM databases which have been edited by multiple contributors. We manually sampled 200 features with $V \geq 15$ from the United Kingdom dataset. These features exhibited a number of interesting characteristics including: long editing timespan from feature creation to current version timestamp,

Table 1: The distribution of version numbers of ways in the four OpenStreetMap datasets.

Versions	**Ireland**	**United Kingdom**	**Germany**	**Austria**
1	$139,722$	$2,106,647$	$6,442,209$	$682,155$
2	$51,306$	$729,014$	$2,126,222$	$178,694$
3	$23,806$	$305,536$	$988,126$	$85,113$
4	$11,274$	$158,557$	$552,740$	$71,552$
5	$6,599$	$84,889$	$336,239$	$67,489$
$6 \rightarrow 10$	$9,296$	$131,836$	$594,210$	$113,652$
$11 \rightarrow 15$	$1,571$	$22,929$	$124,019$	$13,833$
$16 \rightarrow 20$	369	$6,242$	$36,596$	$3,697$
$21 \rightarrow 30$	198	$3,183$	$19,321$	$2,050$
$31 \rightarrow 40$	36	670	$4,294$	520
> 40	15	328	$2,332$	290
Total	$244,192$	$3,549,831$	$11,226,308$	$1,219,045$

multiple unique contributors adding/deleting nodes and tags on the feature, and contributors returning after a number of edits have been made by other contributors. As evident from Table 1 there are very few (relatively speaking) “heavily edited objects” in the OSM database. Yet they offer, in our opinion, the most interesting opportunities for analysis of the collaborative aspect of OSM editing and contribution. The equivalent object in Wikipedia: the “featured articles” is similar in terms of occurrence. In September 2011 there are almost four million articles in the English language Wikipedia but only 3,377 featured articles - which roughly translates to one featured article for every 1000 articles. In our case study databases there are approximately 16 million objects - Just over 12 million of these (about 75%) have only one or 2 versions. Consequently we felt that the choice of $V \geq 15$ was appropriate because of the small number of objects available.

In the Ireland dataset there are 776 features with V or more versions, in the United Kingdom dataset there are $12,804$, in the Germany there are $76,355$, and in the Austrian dataset there are 7950. Total number of features with versions $V \geq 15$ is $97,885$. We randomly selected $25,000$ of these features and finally 653 features were selected from the Ireland dataset, $10,040$ from the United Kingdom dataset, $10,604$ from the German dataset, and $3,367$ from the Austrian dataset. Each dataset is a mix of landuse, highway, amenity, waterways, and natural features. In total the OSM-XML processing and OSM data download from the OSM-API took 920 hours. The scripts to automatically download the OSM-XML using the OSM-API were carefully monitored as connection breaks occurred frequently. We were also mindful of bandwidth limiting on the OSM servers. The scripts were usually only run during normal working hours. The datasets were downloaded and processed during May and June 2011. The data was stored in a PostGIS database.

For each object P the history of edits is downloaded as an OSM-XML file from the OSM API. Suppose that the object P has n versions ($n \geq V$) where $i = 0$ is the first version and $i = n - 1$ is the final or current version. Then each version P_i of P is stored as the tuple represented in Equation 1.

Table 2: The percentage of ways in the four OpenStreetMap datasets with the specified number of unique tags over each object's history.

UniqueTags	**UK**$\%$	**Germany**$\%$	**Austria**$\%$	**Ireland**$\%$
1,2	15.96	35.01	13.51	20.31
3,4	36.11	35.46	21.09	36.92
5,6	27.15	16.94	16.31	23.69
7,8	13.07	6.61	11.7	11.23
9,10	4.68	2.58	8.61	5.38
11,12	1.72	1.52	11.4	2
13,14	0.74	0.84	9.21	0.46
≥ 15	0.56	1.05	8.18	0
Total	10,040	10,604	3,367	653

$$P_i = \Big(u_i, v_i, N_i, \tau_i, c_i, NSR_i, G(i), A(i), L(i), D_i, T(i)\Big) \tag{1}$$

Where the elements of the tuple P_i are as defined as follows: u_i is the user id of the OSM contributor who edited version v_i, v_i is the version of the OSM object, N_i is the number of nodes in object P_i, τ_i is the timestamp for the edit, c_i is the changeset that the edit was saved in, NSR_i is the number of nodes which "survived" from the previous version P_{i-1} of polygon P_i, $G(i)$ is the geometry of P_i, $A(i)$ is the area of $G(i)$ in hectares (only calculated for polygons), $L(i)$ is the length or perimeter of $G(i)$ in meters, D_i is the mean spacing in meters between the adjacent nodes of P_i, and $T(i)$ is the set of tags (keys,values) assigned to this version of P_i which are stored as a comma-separated list. Finally, if specified for each version P_i of the object P a vector data file representation is written out to disk. There are a number of possible output formats: ESRI Shapefile, KML file, or GPX format. This allows for quick visualization within most desktop GIS software and some web-based GIS.

4 Experimental Analysis

In this section we outline results from our experimental analysis of the $25,000$ OSM features. The analysis focuses on three aspects of annotation of these features in OSM namely: (section 4.1) assignment of values to attributes (or in OSM terminology values to tags or tag keys), (section 4.2) types of contribution by the OSM volunteers on these features, and (section 4.3) use of the OSM Map Features page as a controlled vocabulary.

4.1 Tag assignment

One of the major concerns about OSM is that the flexibility of the tagging/annotation model is such that spurious data or noise will be created (Mooney et al., 2010*b*). In table 2 we show the distribution of unique tags (key-value pairs) assigned to the objects in the four OSM datasets. The first column shows the number of unique tags. All values are percentages of the total number of objects in the corresponding OSM dataset (outlined in the final row of the table). A low number (≤ 4) of unique

Table 3: Number of unique values assigned to the “name” tag of objects in the four OSM databases.

Number Names	**UK**	**Ireland**	**Austria**	**Germany**
1	5528 (76.6%)	299 (69.9%)	1804 (70.2%)	3950 (78.1%)
2	1333 (18.5%)	101 (23.6%)	587 (22.8%)	851 (16.8%)
3	280 (3.9%)	19 (4.4%)	148 (5.8%)	195 (3.9%)
4	58 (0.8%)	6 (1.4%)	27 (1.1%)	49 (1%)
≥ 5	15 (0.2%)	3 (0.7%)	5 (0.2%)	15 (0.3%)

tags can indicate stable tags which remain unchanged over the lifetime of the object. A higher number of unique tags can reveal either: a more detailed set of tags or frequent changes to the values of tags over the lifetime of the object (referred to as tag “flip-flopping”). There are some interesting observations. In the case of Austria one can notice the large percentage of objects with ≥ 9 unique tags. We speculate that this could be caused by the large bulk import of government spatial data into the Austrian OSM database with very rich metadata. Germany has the highest percentage of objects in the four OSM datasets with just 1 or 2 unique tags. In the four datasets some objects were available without any tags assigned. Often this problem was corrected very quickly (within one hour). However, in Austria 47 objects, in UK 310 objects, in Ireland 46 objects, and Germany 398 objects had an empty tag set for at least one day.

Table 3 summarises the number of unique values assigned to the “name” tag of objects in the four OSM databases. As expected 70% or more of the objects that have a “name” attribute with an assigned value which remains unchanged up and until the current version. The high percentage of objects having 2 name value assignments is probably a result of: placename spelling errors, incorrect naming, or the splitting of a single polygon or way into two or more new objects. From our knowledge of the data we believe that contributor disagreement, spelling errors, and uncertainty in local knowledge (possibly resulting from aerial imagery tracing rather than physical sampling) are responsible for the assignment of 3 or more values to the “name” tag of any object. For the purpose illustration we provide tables 4, 5, and 6 outlining the edit history of three road polylines in OSM where there are changes to either the road “name” (table 4 and table 6) or the highway designation attribute (table 5). The aesterick symbol indicates the current version of the objects in the OSM database. Each table shows the date of edit, the version number, the tag value, and the ID of the contributor who made the edit. In all three cases multiple contributors are involved. Some edits made on the same day probably, in our opinion, correspond to self-corrections by the contributors who made them.

Table 7 is similar to the results presented in Table 3. In this table we summarise the number of unique values that the “highway” attribute is assigned for all objects in the four databases with the “highway” tag. There are some interesting observations. A very small, but not negligible, number of objects have a high number of changes of highway designation. For example object 9779683 in Germany has 4 different values assigned to it’s highway attribute over it’s 29 version history. These values are ‘primary’, ‘tertiary’, ‘residential’, ‘secondary’. There are 7 unique contributors to this object. A significant percentage of objects in all four databases have 3 or more unique values assigned to their highway attribute. We believe that it is unlikely that real-world physical highways (motorways, roads, paths, etc) could change their designation this frequently. For example: UK (9.3%), Ireland (11.8%), Austria (16.7%), and Germany (15.1%). With comparison to a ground-truth dataset it is difficult

Table 4: Example of changes to the value of the name attribute of the road polyline 24276789 (England)

Date of Edit	**Version**	**Name**	**Contributor**
2008-05-08	1	Oakthorp Drive	35691
2008-05-09	6	Over Green Drive	35691
2008-05-09	9	Oak Thorp Cr	35691
2008-05-09	10	Oak Thorp Dr	35691
2008-05-11	14	Oak Thorp Dr; Broomcroft Rd	35691
2008-05-11	15	Oak Thorp Dr	35691
2010-02-07	18	Oak Thorp Drive	9065
*2010-08-24	19	Oak Thorpe Drive	35691

Table 5: Example of changes to the value of the name attribute of the road polyline 9779683 (Germany)

Date of Edit	**Version**	**Highway**	**Contributor**
2007-10-18	1	primary	16631
2007-10-18	2	tertiary	16631
2007-12-06	8	tertiary; primary	16631
2007-12-06	9	residential	16631
2008-01-04	10	residential; secondary	16631
2008-01-04	14	residential	16631
2008-06-20	22	tertiary	46829
*2010-03-21	29	tertiary	95223

to precisely understand the reason for the tag “flip-flopping” with the highway attribute. We believe this could demonstrate uncertainty amongst contributors regarding the designation for a given highway object. This could also reveal a deeper issue of semantics within the OSM Map Features. Different contributors may have conflicting understand of similar descriptions such as “living-street” and “residential”.

4.2 Influence of Contributors

In our case study datasets there are, 2779 unique contributors to the UK dataset, 355 to the Ireland dataset, $1,485$ in Austria, and $9,325$ for the Germany dataset. Any contributor to OSM can add tags or edit existing tags on OSM objects, regardless if they created them or not. Haklay et al. (2010) and Girres and Touya (2010) show that increases in the number of OSM contributors in an area is strongly related to an increase in geometric data quality and spatial data volume. What effects do the number of contributors for each object have on the number of changes to the “name” tag or changes to the designation value of “highway” attributes in table 3 and table 7? We calculated the correlation and the Spearman correlation (ρ and p-value respectively) for the number of unique contributors to each object against the number of tag “flip-flops” on the “highway” tag. Objects are included if they had been assigned a “highway” tag for more than $V/2$ versions of their history. Unfortunately the results are inconclusive.

Table 6: Example of changes to the value of the name attribute of the road polyline 4755815 (Scotland)

Date of Edit	**Version**	**Name**	**Contributor**
2007-06-14	1	A199	6871
2008-01-24	2	null	5121
2009-03-18	17	Edinburgh Road	108345
2011-01-04	24	Milton Road East	364126
2011-01-04	25	Edinburgh Road	364126
2011-01-13	27	Milton Road East	364126
2011-01-13	28	Edinburgh Road	364126
*2011-02-10	29	Edinburgh Road	108345

Table 7: Number of unique values assigned to the “highway” tag of objects in the four OSM databases. The column ‘highway’ indicates the number of unique values assigned.

Highway	**UK**	**Ireland**	**Austria**	**Germany**
1	4999 (59.4%)	298 (50.5%)	1110 (47.1%)	495 (54.8%)
2	2621 (31.2%)	222 (37.6%)	855 (36.3%)	271 (30%)
3	650 (7.7%)	60 (10.2%)	305 (12.9%)	110 (12.2%)
4	117 (1.4%)	8 (1.4%)	78 (3.3%)	22 (2.4%)
≥ 5	22 (0.3%)	2 (0.3%)	10 (0.4%)	5 (0.6%)

We calculated the two-sided p-value for a hypothesis test where the null hypothesis was that the number of contributors and the number of tag "flip-flops" were uncorrelated. A p-value exceeding 0.05 corresponded to accepting the null hypothesis. The results were as follows $(N, corr, \rho, p - value)$: UK $(8210, 0.21, 0.13, 0.561)$, Ireland $(590, -0.18, -0.07, 0.46)$, Austria $(2350, 0.09, 0.08, 0.061)$, and Germany $(903, 0.22, 0.18, 0.112)$. The results are disappointing but expected. The correlation values in all cases correspond to very weak correlations. Similar results were calculated for tag "flip-flops" on the "name" tag. While the correlation values are weak to moderate no conclusion can be drawn to indicate a relationship between number of contributors to an object and tag "flip-flopping" on the object. We calculated the correlation between the number of unique contributors to an object and the number of tags at the current version. The results did not reveal any obvious relationship. Correlations were: Germany: 0.05, UK: 0.19, Ireland: 0.18, and Austria: 0.12. On the one hand it is a valid assumption to assume the number of tags will increase as more contributors are involved in editing an object. However, Kessler et al. (2011) state that no changes to tagging over many versions, under the eyes of many contributors, could be used as a mechanism for assigning trust or stability to an object's tags.

4.3 Adherence to OSM controlled vocabulary

As discussed above the OSM Map Features (OSM Map Features, 2011) page provides a listing of the most popular values for the most frequently occurring attributes (highway, amenity, landuse, natural, etc). Interestingly we found that there were a core set of values causing non-compliance. For example incorrect spelling of "landuse=forest" as "forrest,forestry,forrestry" while "highway=residential" had incorrect spellings of "ressidential,resident,residental". Errors such as these could potentially be automatically corrected. For the "highway" attribute there are 37 core values (primary, motorway, cycleway, livingstreet, etc). For the "landuse" attribute there are 29 core values (forest, farmyard, industrial, grass, etc). While editor software for OSM usually present these core values in drop-down-list selection functionality contributors can type these values in as free-text or supply their own values for the attribute. For example, in the UK there are 577 objects which have the "landuse" attribute at some stage of their history. In total 39 values were assigned to "landuse" attribute tags (so 39 - $29 = 10$ free text tags not defined in OSM Map Features. From a visual inspection of these spelling errors accounted for the majority of these variants. In Table 9 a summary of the number of objects with landuse or highway attributes (at 1st and current version) is provided. The number of objects with these attributes is shown in the *Objects* column. The number of these objects where the values for either landuse or highway attributes are drawn from the Map Features controlled vocabulary is shown in the *Compliance* column. In all cases the number of objects with these tags increases from the first to the current version. Compliance with the Map Features controlled vocabulary is very good overall. Being compliant with the map features controlled vocabulary does not in any way indicate that this attribute assignment is currently correct and would need to be confirmed by comparison to ground-truth datasets.

5 Conclusions and Future Work

In this paper we have investigated how spatial objects are tagged in OSM databases. Four OSM databases we selected and from these $25,000$ heavily edited objects were chosen for analysis. The paper begun with an introduction to "tagging" in OSM. This was followed by a comprehensive overview of the literature on OSM. We then described the process of choosing heavily edited

Table 8: Overall usage of values from the OSM controlled vocabulary ’Map Features’ from all unique values assigned to “landuse” and “highway” tags. The compliance column indicates the number of unique values found with the number of these *not* in the controlled vocabulary in brackets.

Database	**Attribute**	**Compliance**	**Observations**
UK	Landuse	39 (10)	Spelling Errors
UK	Highway	138 (101)	Spelling errors ’pedestrianissed’, ’tersiary’ and assigning the name of the road or highway to the highway tag
Ireland	Landuse	5 (0)	All valid
Ireland	Highway	30 (0)	All valid
Germany	Landuse	105 (76)	Spelling errors ’medow’, ’forrest’, and invalid assignments ’fruit trees’
Germany	Highway	49 (12)	Street names assigned to highway attribute
Austria	Landuse	72 (43)	Spelling errors of core values, invalid values
Austria	Highway	118 (81)	Spelling errors, multiple value assignments, alternative values from bulk import

Table 9: Compliance of tagging with the map features controlled vocabulary - first and current versions.

Database	**Attribute**	**Version**	**#Objects**	**#Compliant**
UK	Highway	1^{st}	6730	6650
UK	Highway	Current	8269	8267
Ireland	Highway	1^{st}	456	455
Ireland	Highway	Current	579	579
Austria	Highway	1^{st}	2019	1760
Austria	Highway	Current	2326	2325
Germany	Highway	1^{st}	608	587
Germany	Highway	Current	697	697
UK	Landuse	1^{st}	463	437
UK	Landuse	Current	577	558
Ireland	Landuse	1^{st}	6	6
Ireland	Landuse	Current	7	7
Austria	Landuse	1^{st}	516	406
Austria	Landuse	Current	914	910
Germany	Landuse	1^{st}	5253	5154
Germany	Landuse	Current	7058	7039

objects and working with OSM-XML data. The locations of Ireland, United Kingdom, Germany, and Austria were chosen because of the home location of the authors and that the activity of the OSM communities in the other three regions. The analysis could be easily extended to other regions. Table 1 shows that over 90% of objects in the four OSM databases have ≤ 3 versions. This makes it difficult to undertake analysis to investigate collaborative editing on these objects. Subsequently our analysis choose to investigate "heavily edited" objects. These offer a similar concept to the Wikipedia Featured Article. Heavily edited articles in Wikipedia are usually those that gain the status of "featured article". Featured articles are recognized as articles of high quality, with a long history of collaborative editing, and have become relatively stable (no major recent edits) (Anderka et al., 2011). Welser et al. (2011) explain that usually heavily edited articles in Wikipedia gain the status of "featured article" and are subsequently recognized as articles of high quality. Korfiatis et al. (2006) based their analysis of quality of Wikipedia articles on successive edits and therefore focused on articles with a long edit history. Hecht and Gergle (2010) focus on articles which have been edited frequently, particularly those by the same contributor. Nemoto et al. (2011) indicates that quality increases, and stabilizes, the more contributors work on a given article.

The tagging and annotation of these heavily edited objects in OSM is restricted to the use of a small number of tags. In all four datasets at least 50% of objects use 6 tags or less over their history. We found the use of tags such as "source" and "description" (to indicate how the data was captured etc.) was sparse. Only 3.5% of the $25,000$ objects used one or both of these tags. Tag "flip-flopping" occurs where the values assigned to tags such as "name" and "highway" change multiple times. Table 3 and table 7 show that a small, but not negligible, percentage of objects have their "name" or "highway" tags assigned different values over the object lifetime. The OSM Map Features page offers a controlled vocabulary from which contributors can choose values for tags such as "landuse" and "highway". Table 9 shows the number of objects which draw the values for their "landuse" and "highway" tags from the Map Features controlled vocabulary. The rate of compliance is very high ($> 98\%$ for the current versions of all objects). However this compliance does not imply that the current values assigned to these attributes are correct. Table 8 shows wide variations on the controlled vocabulary are used with table 7 and table 3 showing significant "flip-flopping" of values assigned to the "highway" and "name" tags. Finally, no relationship was found to exist in our four datasets between the number of contributors to an object and the number of tags or tag "flip-flopping" on that object. Overall, this work shows that there are issues in how contributors tag and annotate spatial features in OSM. These issues need to be addressed before OSM can be considered for use in "serious geomatics applications" (Mooney et al., 2010*b*; Over et al., 2010)

As described in section 3.2 our database of history for the selected $25,000$ objects is a very detailed record of contributor activity to OSM over a period of approximately four years. This provides us with a very rich dataset from which future research work can be developed. Haklay et al. (2010) and Girres and Touya (2010) show that increases in the number of OSM contributors in an area are strongly related to an increase in geometric data quality and spatial data volume. An immediate issue for future work would be comparison of the tags of these $25,000$ with ground-truth data to investigate if a relationship exists between the number of contributors and accurate of tagging. There are no moderators for content in OSM. Contributors can take a 'moderator' responsibility for a particular OSM region or a set of objects. It would be interesting to conduct a survey of OSM contributors to understand the causes of tag value changes for example. This would allow us to formulate some indication of the methods of contributions of different communities of OSM from different countries etc. This work could also include an analysis of the geometric and positional accuracy of heavily edited objects over time measured against some ground truth dataset. As outlined in section 3.3. there are just over 97,000 objects in the four databases with V $\geq$ 15. From the complete set of objects in all four databases this represents less than 1% of all objects. For future work we will also consider reducing the threshold value V to investigate the effects it has on our analysis and results. Finally, in the field of visualisation we feel there is scope for work on the visualisation of the historical evolution of features in OSM. Gilbert and Karahalios (2009) remark that they "see vast potential for social visualization to make large impacts on social production projects" because added value can be gained for both those involved in the project and outside it from being able to "to observe the long-term impacts on motivation and production in real, working social production communities". Several authors (Suh et al., 2008; Pirolli et al., 2009; Riche et al., 2010) claim that the increased exposure to the editing/authoring histories, using visualisation software applications, for collaborative knowledge projects like Wikipedia, increases credibility judgements and offers transparency. This can eventually lead to "improvements in the interpretation, communication, and trustworthiness" (Suh et al., 2008) of collaboratively generated knowledge. We feel that this could extend to include OpenStreetMap and other VGI projects.

References

Anderka, M., Stein, B. and Lipka, N. (2011), Towards automatic quality assurance in wikipedia, *in* 'Proceedings of the 20th international conference companion on World wide web', WWW '11, ACM, New York, NY, USA, pp. 5–6.

Antin, J. (2011), My kind of people?: perceptions about wikipedia contributors and their motivations, *in* 'Proceedings of the 2011 annual conference on Human factors in computing systems', CHI '11, ACM, New York, NY, USA, pp. 3411–3420.

Ather, A. (2009), "A quality analysis of openstreetmap data", *Quality* , p. 81.

Ballatore, A. and Bertolotto, M. (2011), Semantically enriching vgi in support of implicit feedback analysis, *in* K. Tanaka, P. Fröhlich and K.-S. Kim, eds, 'Web and Wireless Geographical Information Systems', Vol. 6574 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 78–93.

Brando, C. and Bucher, B. (2010), Quality in user generated spatial content: A matter of specifications, *in* M. Painho, M. Y. Santos and H. Pundt, eds, 'Proceedings AGILE 2010: The 13th AGILE International Conference on Geographic Information Science', Springer Verlag, Guimarães, Portugal.

Budhathoki, N., Bruce, B. and Nedovic-Budic, Z. (2008), "Reconceptualizing the role of the user of spatial data infrastructure", *GeoJournal* , Vol. 72, Springer Netherlands, pp. 149–160. 10.1007/s10708-008-9189-x.
URL: *http://dx.doi.org/10.1007/s10708-008-9189-x*

Bulterman, D. C. A. (2004), "Is it time for a moratorium on metadata?", *IEEE MultiMedia* , Vol. 11, pp. 10–17.

Cantador, I., Konstas, I. and Jose, J. M. (2011), "Categorising social tags to improve folksonomy-based recommendations", *Web Semantics: Science, Services and Agents on the World Wide Web* , Vol. 9, pp. 1 – 15.

Ciepluch, B., Mooney, P., Jacob, R. and Winstanley, A. C. (2009), "Using openstreetmap to deliver location-based environmental information in ireland", *SIGSPATIAL Special* , Vol. 1, ACM, New York, NY, USA, pp. 17–22.

Coleman, D. J., Georgiadou, Y. and Labonte, J. (2009), "Volunteered geographic information : The nature and motivation of produsers", *International Journal of Spatial Data Infrastructures Research* , Vol. 4, pp. 332–358.

Corcoran, P. and Mooney, P. (2011), Topologically consistent selective progressive transmission, *in* S. Geertman, W. Reinhardt and F. Toppen, eds, 'Advancing Geoinformation Science for a Changing World', Vol. 1 of *Lecture Notes in Geoinformation and Cartography*, Springer

Berlin Heidelberg, pp. 519–538.

Corcoran, P., Mooney, P. and Winstanley, A. C. (2011), “Planar and non-planar topologically consistent vector map simplification”, *International Journal of Geographic Information Science.* , Vol. x, Taylor & Francis, Inc., Bristol, PA, USA, p. to appear.

Flanagin, A. J. and Metzger, M. J. (2008), “The credibility of volunteered geographic information”, *GeoJournal* , Vol. 72, pp. 137–148.

Geofabrik (2011*a*), ‘Data downloads for openstreetmap data’, Online service: `http://www.geofabrik.de/data/download.html`.

Geofabrik (2011*b*), ‘Mapcompare - visual comparision of google maps and openstreetmap’, Geofabrik website - `http://tools.geofabrik.de/mc/` - checked July 2011.

Gilbert, E. and Karahalios, K. (2009), “Using social visualization to motivate social production”, *IEEE Transactions on Multimedia* , Vol. 11, pp. 413 –421.

Girres, J.-F. and Touya, G. (2010), “Quality assessment of the french openstreetmap dataset”, *Transactions in GIS* , Vol. 14, p. 435459.

Goodchild, M. (2007), “Citizens as sensors: the world of volunteered geography”, *GeoJournal* , Vol. 69, Springer Netherlands, pp. 211–221. 10.1007/s10708-007-9111-y.

Goodchild, M. (2008), “Commentary: whither vgi?”, *GeoJournal* , Vol. 72, Springer Netherlands, pp. 239–244. 10.1007/s10708-008-9190-4.

Haklay, M. (2010), “How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets”, *Environment and Planning B: Planning and Design* , Vol. 37, pp. 682–703.

Haklay, M., Basiouka, S., Antoniou, V. and Ather, A. (2010), “How many volunteers does it take to map an area well? the validity of linus’ law to volunteered geographic information”, *The Cartographic Journal* , Vol. 47, pp. 315 – 322.

Hecht, B. J. and Gergle, D. (2010), On the ”localness” of user-generated content, *in* ‘Proceedings of the 2010 ACM conference on Computer supported cooperative work’, CSCW ’10, ACM, New York, NY, USA, pp. 229–232.

URL: *http://doi.acm.org/10.1145/1718918.1718962*

Jacob, R., Mooney, P., Corcoran, P. and Winstanley, A. C. (2010), “Haptic-gis: exploring the possibilities”, *SIGSPATIAL Special* , Vol. 2, ACM, New York, NY, USA, pp. 13–18.

Javanmardi, S. and Lopes, C. (2010), Statistical measure of quality in wikipedia, *in* ‘Proceedings of the First Workshop on Social Media Analytics’, SOMA ’10, ACM, New York, NY, USA, pp. 132–138.

Kessler, C., Janowicz, K. and Bishr, M. (2009), An agenda for the next generation gazetteer: geographic information contribution and retrieval, *in* ‘Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems’, GIS ’09, ACM, New York, NY, USA, pp. 91–100.

Kessler, C., Trame, J. and Kauppinen, T. (2011), Tracking editing processes in volunteered geographic information: The case of openstreetmap, *in* M. Duckham, A. Galton and M. Worboys, eds, ‘Proceedings of Workshop on Identifying Objects, Processes and Events in Spatio-Temporally Distributed Data (IOPE)’, Conference on Spatial Information Theory 2011 (COSIT’11), p. to appear.

Korfiatis, N., Poulos, M., and Bokos, G. (2006), “Evaluating authoritative sources using social networks: an insight from wikipedia”, *Online Information Review* , Vol. 30, pp. 252–262.

Langläufer (2011), ‘The osm history browser web application’, Online Service: `http://osm.virtuelle-loipe.de/history/`.

Liu, D., Wang, M., Hua, X.-S. and Zhang, H.-J. (2011), “Semi-automatic tagging of photo albums via exemplar selection and tag inference”, *IEEE Transactions on Multimedia* , Vol. 13, pp. 82 –91.

Ludwig, I., Voss, A. and Krause-Traudes, M. (2011), A comparison of the street networks of navteq and osm in germany, *in* S. Geertman, W. Reinhardt and F. Toppen, eds, ‘Advancing Geoinformation Science for a Changing World’, Vol. 1 of *Lecture Notes in Geoinformation and Cartography*, Springer Berlin Heidelberg, pp. 65–84.

Mooney, P. and Corcoran, P. (2011*a*), Accessing the history of objects in openstreetmap, *in* S. Geertman, W. Reinhardt and F. Toppen, eds, 'Proceedings AGILE 2011: The 14^{th} AGILE International Conference on Geographic Information Science', Springer Verlag, Utrecht, The Netherlands, p. p155.

Mooney, P. and Corcoran, P. (2011*b*), Annotating spatial features in openstreetmap, *in* 'Proceedings of the 19th annual gis research uk (gisruk) 2011', Portsmouth, England.

Mooney, P. and Corcoran, P. (2011*c*), Can volunteered geographic information be a participant in eenvironment and sdi?, *in* J. Hrebícek, G. Schimak and R. Denzer, eds, 'Environmental Software Systems. Frameworks of eEnvironment', Vol. 359 of *IFIP Advances in Information and Communication Technology*, Springer Boston, pp. 115–122.

Mooney, P., Corcoran, P. and Winstanley, A. C. (2010*a*), A study of data representation of natural features in openstreetmap, *in* 'Proceedings of the 6th GIScience International Conference on Geographic Information Science', GIScience 2010, University of Zurich, p. p150.

Mooney, P., Corcoran, P. and Winstanley, A. C. (2010*b*), Towards quality metrics for openstreetmap, *in* 'Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems', GIS '10, ACM, New York, NY, USA, pp. 514–517.

Nemoto, K., Gloor, P. and Laubacher, R. (2011), Social capital increases efficiency of collaboration among wikipedia editors, *in* 'Proceedings of the 22nd ACM conference on Hypertext and hypermedia', HT '11, ACM, New York, NY, USA, pp. 231–240.
URL: *http://doi.acm.org/10.1145/1995966.1995997*

OpenStreetMap (2011*a*), 'The map-features page on openstreetmap - a guide to the recommended tagging of osm features', Online service: `http://wiki.openstreetmap.org/wiki/Map_Features`.

OpenStreetMap (2011*b*), 'Openstreetmap automated import and editing code of conduct', Online Wiki Page: `http://wiki.openstreetmap.org/wiki/Automated_Edits/Code_of_Conduct`.

OSM API (2011), 'Openstreetmap editing api for fetching and saving raw geodata from/to the global openstreetmap database', Online Wiki Page: `http://wiki.openstreetmap.org/wiki/API`.

OSM History Viewer (2011), 'The osm changeset visualiser web application', Online Service: `http://osmhv.openstreetmap.de/index.jsp`.

OSM Map Features (2011), 'The map features page guide to tagging features on openstreetmap', Online Wiki Page: `http://wiki.openstreetmap.org/wiki/Map_Features` (checked July).

OSM-Stats (2011), 'The openstreetmap statistics page', Online - dynamically updating - `http://www.openstreetmap.org/stats/data_stats.html`.

Over, M., Schilling, A., Neubauer, S. and Zipf, A. (2010), "Generating web-based 3d city models from openstreetmap: The current situation in germany", *Computers Environment and Urban Systems* , Vol. 34, Elsevier Ltd, pp. 496–507.

Pirolli, P., Wollny, E. and Suh, B. (2009), So you know you're getting the best possible information: a tool that increases wikipedia credibility, *in* 'Proceedings of the 27th international conference on Human factors in computing systems', CHI '09, ACM, New York, NY, USA, pp. 1505–1508.

Prasarnphanich, P. and Wagner, C. (2011), "Explaining the sustainability of digital ecosystems based on the wiki model through critical-mass theory", *IEEE Transactions on Industrial Electronics* , Vol. 58, pp. 2065 –2072.

Qian, X., Di, L., Li, D., Li, P., Shi, L. and Cai, L. (2009), Data cleaning approaches in web2.0 vgi application, *in* 'Geoinformatics, 2009 17th International Conference on', pp. 1 –4.

Ramm, F. (2011), 'Displaying historic openstreetmap coverage', Online at Geofabrik Labs: `http://labs.geofabrik.de/history/`.

Riche, N. H., Lee, B. and Chevalier, F. (2010), ichase: supporting exploration and awareness of editing activities on wikipedia, *in* 'Proceedings of the International Conference on Advanced Visual Interfaces’, AVI ’10, ACM, New York, NY, USA, pp. 59–66.

Roick, O. (2011), ‘OSMatrix - visualisation of contributions to osm’, Web Service from Institute of Geography Heidleberg: Online at `http://koenigstuhl.geog.uni-heidelberg.de/osmatrix/` - checked April 2011.

Smart, P. D., Quinn, J. A. and Jones, C. B. (2011), “City model enrichment”, *ISPRS Journal of Photogrammetry and Remote Sensing* , Vol. 66, pp. 223–234. Quality, Scale and Analysis Aspects of Urban City Models.

Stein, K. and Hess, C. (2007), Does it matter who contributes: a study on featured articles in the german wikipedia, *in* ‘Proceedings of the eighteenth conference on Hypertext and hypermedia’, HT ’07, ACM, New York, NY, USA, pp. 171–174.

Suh, B., Chi, E. H., Kittur, A. and Pendleton, B. A. (2008), Lifting the veil: improving accountability and social transparency in wikipedia with wikidashboard, *in* ‘Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems’, CHI ’08, ACM, New York, NY, USA, pp. 1037–1040.

Trame, J. and Kessler, C. (2011), Exploring the lineage of volunteered geographic information with heat maps., *in* ‘GeoViz 2011 - Linking Geovisualization with Spatial Analysis and Modelling’.

van Exel, M., Dias, E. and Fruijtier, S. (2010), The impact of crowdsourcing on spatial data quality indicators, *in* ‘Proceedings of the 6th GIScience International Conference on Geographic Information Science’, GIScience 2010, University of Zurich, p. p213.

Wallgrün, J. O., Wolter, D. and Richter, K.-F. (2010), Qualitative matching of spatial information, *in* ‘Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems’, GIS ’10, ACM, New York, NY, USA, pp. 300–309.

Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G. and Smith, M. (2011), Finding social roles in wikipedia, *in* ‘Proceedings of the 2011 iConference’, iConference ’11, ACM, New York, NY, USA, pp. 122–129.

Zielstra, D. and Hochmair, H. H. (2011), “Digital street data - free versus proprietary”, *GIM International* , Vol. 25, pp. 28–06–2011.

Zielstra, D. and Zipf, A. (2010*a*), A comparative study of proprietary geodata and volunteered geographic information for germany, *in* M. Painho, M. Y. Santos and H. Pundt, eds, ‘Proceedings AGILE 2010: The 13th AGILE International Conference on Geographic Information Science’, Springer Verlag, Guimarães, Portugal.

Zielstra, D. and Zipf, A. (2010*b*), Quantiative studies on the data quality of openstreetmap in germany, *in* R. Weibel and S. Fabrikant, eds, ‘Proceedings GIScience 2010: The Sixth International Conference on Geographic Information Science’, Springer Verlag LNCS, Zurich, Switzerland, pp. 20–26.