Assessing the Completeness of Bicycle Trail and Designated Lane Features in OpenStreetMap for the United States and Europe

Revised version submitted: November 15, 2012
Word Count: 5415 words plus 8 figures/tables at 250 words each = 7415 words

Hartwig H. Hochmair
University of Florida, Geomatics Program
3205 College Avenue
Ft. Lauderdale, FL-33314
Phone: (954) 577-6317
Fax: (954) 475-4125
E-mail: hhhochmair@ufl.edu

Dennis Zielstra
University of Florida, Geomatics Program
3205 College Avenue
Ft. Lauderdale, FL-33314
Phone: (954) 577-6392
E-mail: dzielstra@ufl.edu

Pascal Neis
University of Heidelberg, Geoinformatics Research Group
Berliner Str. 48
Heidelberg
D-69120 Germany
Phone: 01149-6221-54-5504
E-mail: neis@uni-heidelberg.de

ABSTRACT

This paper assesses the completeness of OpenStreetMap data for cycling features, in particular bicycle trails and designated lanes, for selected cities in the United States and Europe. While most available comprehensive road network datasets, either from commercial data vendors or public agencies, are tailored towards motorized traffic, OpenStreetMap as a community based, open access platform for geospatial vector data provides a viable alternative for data retrieval of cycling feature data. The analysis in this study reveals a steady growth of cycling related data in the OpenStreetMap platform over the past few years, rendering the data more complete and appropriate to be used as base map for planning studies on non-motorized transportation. A comparison with the Google Maps Bicycling layer shows that the data quality of OpenStreetMap designated lanes is particularly high. OpenStreetMap bicycle trail data are somewhat more erroneous through missing data and incorrectly classified trails, however still of relatively good quality. For practical purposes it is recommended to check OpenStreetMap trail data against the Google Maps Bicycling layer before an analysis is conducted based on OpenStreetMap trail data.

INTRODUCTION

Digital data repositories of bicycle infrastructure in a transportation network, such as bicycle lanes or off-street paths, provide an important basis for a variety of bicycle transportation planning and analysis tasks, including latent demand analysis (*9, 1, 6*), bicycle trip planning applications (*7, 18*), and analysis of bicycle travel behavior (*2, 8*). A variety of public and proprietary road network data exist which can be used for network analysis regarding motorized traffic. Examples of publicly administered datasets are the freely available TIGER/Line data from the U.S. Census Bureau and the ATKIS (Amtliches Topographisch-Kartographisches Informationssystem) proprietary data set for Germany hosted by the German Federal Office of Cartography and Geodesy. Examples of commercial data providers are NAVTEQ and TomTom whose data are often used in car navigation systems, and recently also for pedestrian navigation in- and outdoors (e.g., NAVTEQ Destination Maps or NAVTEQ Discover Cities). However, none of these data sources considers bicycle related network data, such as cycling facilities or bicycle trails. Transportation planning tasks at the local level that involve non-motorized traffic, in particular bicycle transportation mode, are therefore usually based on data collected by different local stakeholders, e.g. county and city planning departments, and therefore fragmented and not easily accessible to the public. OpenStreetMap (OSM), which is an open, freely available international repository of geographic data, provides a feasible alternative to scattered, locally stored cycling data. OSM uses widely accepted, since community developed, coding conventions for bicycle infrastructure in a transportation network and hosts numerous transportation feature layers in a publicly accessible and editable database, including bicycle facilities. The OSM data volume is rapidly growing, and new data tagging proposals from OSM data contributors are underway for more detailed mapping of bicycle facilities, making OSM a promising repository for the cyclist community. One major advantage of OSM is that it provides a relatively homogeneous tagging structure of spatial data for the whole world, whereas data sets by local agencies often vary in classification of facilities and attribute schemes, which causes a problem of interoperability for comprehensive data analysis and re-use.

The objective of this paper is to analyze the quality, more specifically, the completeness of bicycle related network data in OSM. While road network variables provided by the OSM tagging scheme are not (yet) detailed enough to calculate bicycle level of service measures (*5*), basic attributes and features, such as bike trails and designated lanes on roads, are already available, facilitating a basic level of bicycle related network analysis based on OSM data. The results of this research should help for transportation planners and analysts to understand the pros and cons of OSM data and assist them in deciding whether OSM is a viable data source for their transportation analysis tasks.

OPENSTREETMAP

OpenStreetMap is a collaborative mapping project (available at openstreetmap.org) with data contributed by different users from various backgrounds. The information is collated on a central database and distributed in multiple digital formats through the World Wide Web. For small areas, the data can be retrieved in osm-xml format directly from the OpenStreetMap.org Web site. The entire OSM planet dataset can be downloaded from planet.openstreetmap.org but needs to be processed and imported into a database, such as PostgreSQL. Other Websites, such as Cloudmade.com and Geofabrik.de provide OSM data download for pre-selected administrative areas, such as continent, country, or state, also in shapefile format among others. OSM provides spatial data from a large selection of themes, including roads, transit lines, tourist sites, or land use layers. The rapid growth of the OSM database with its free street data made OSM a hot topic among the geospatial research community over the recent years. One direction of research is concerned with user contribution patterns, i.e., which users contribute what amount of data (*15, 11*). Another research direction focusses on data quality since OSM data are not administered by a regulatory instance but contributed by non-professional individuals with generally little experience or training. Previous studies analyzed various aspects of data quality of OSM street data, including completeness (*12, 19, 20*), positional accuracy (*4*), and attribute accuracy (*12, 10*). A recent novel research approach utilizes the editing history of individual OSM features for data quality assessment (*17*). While data completeness of OSM street data has already been analyzed for motorized transportation and pedestrian navigation in some papers, we are not aware of any study that assessed the completeness of OSM bicycle feature data, such as bicycle trails and designated on-street lanes, which is the objective of this paper.

Tagging of OpenStreetMap Features

OSM employs an open tagging system, which does not enforce a set of standard tags but allows each contributor to use his or her own tagging scheme for mapped features. However, OSM provides guidance for best practices on tagging and recommended tags. The open tagging system makes it necessary to acknowledge that the same feature, e.g. a bicycle trail, can be tagged in different ways by different users. In OSM coding, elements (nodes, ways, and relations) are described through tags. Each tag consists of a key and a value which are free format textual fields. Tags in OSM documentation are written as *Key=value*, where a key broadly describes an element (e.g. a highway) or attribute associated with an element (e.g. speed limit), and the value more specifically describes its accompanying key. OSM uses a total of 26 primary feature keys including *building*, *highway*, or *landuse*. Features can further be annotated with additional tags, such as mode designation of a road. A summary of commonly used tags can be found in the OSM wiki for Map Features (*14*). A comprehensive manual for using and contributing OSM data, with numerous references to the OSM wiki, is a book by Ramm et al. (*16*). A brief introduction to relevant tags for multi-modal transportation, including cycling, is provided in (*5*).

Roads and Footpaths

Roads and footpaths are tagged as *highway=**. Roads with motorized traffic have tag values *motorway, trunk, primary, secondary, tertiary, and residential*. Unpaved roads, such as roads for agricultural use or gravel roads in the forest, are denoted by *highway=track*. A *highway=service* tag is used for service roads to, or within an industrial estate, camp site, business park, car park etc. With respect to bicycle facilities we distinguish in our analysis between bicycle trails and designated lanes. Google Maps uses in its Bicycling layer a similar distinction, including bicycle trails and dedicated lanes. This similarity in feature classification facilitates a comparison in data completeness between OSM and Google Maps data for data quality assessment. We use the term bicycle trail for an off-road path that permits bicycle use, either exclusively, or combined with other non-motorized transportation modes, such as walking. Bicycle trails also include bicycle paths along roads that are physically separated from car traffic, which is referred to as cycle tracks in terms of OSM definitions.

Bicycle Trails and Designated Lanes

There are several options to tag a bicycle trail in OSM. Oftentimes, a bicycle trail is tagged as *highway=cycleway*, indicating that the used way is mainly or exclusively for bicycles. This is also the recommended coding for cycle tracks along roads. In general, *highway=** tags, which are often combined with additional tags, can also be used to express ways with non-motorized traffic that allow cycling. The *highway=path* tag codes a non-specific or shared-use path that is open to the public and not intended for motor vehicles. Specific path designations and access permissions can be expressed through additional tags, such as *bicycle=designated* (designation) or *bicycle*=yes (public has a right of way when traveling on a bicycle). The *bicycle=designated* tag may imply extra usage rights for cycling. But it may also just indicate a suggested bicycle route without any particular cycling facilities (e.g. lanes) that is shared with motorized traffic. The *official* value for an access tag, e.g. *bicycle=official*, indicates a way legally dedicated to specific modes of travel. Similarly *bicycle=yes* expresses permission to use a bicycle. The *highway=footway* tag maps minor paths used mainly or exclusively by pedestrians, and the *highway=pedestrian* tag is used for town centers and civic areas with hard surfaces provided for pedestrians to walk. The *highway=bridleway* tag shows a way intended for use by pedestrians and horse riders, where it can be assumed that cyclists are also permitted unless explicitly prohibited. The *cycleway=track* and *cycleway=opposite_track* tags, which are used in addition to a *highway* key, map a cycle path that is separated from cars. While the *highway=track* tag also permits cars (e.g. agricultural vehicles), the additional *motor_vehicle= no* tag restricts its use to non-motorized traffic.

Designated lanes that are shared with a highway feature are tagged as *highway=* + cycleway=lane*. If a two-way road has a bicycle lane on just one side running in just one direction, the *cycleway:right=lane* or *cycleway:left=lane* tags can be used. The *cycleway=opposite_lane* tag maps a lane where bicycles may go in the direction opposite of other traffic. Cycle lanes shared with bus and taxi lanes are tagged as *cycleway=shared_busway*, *cycleway:left=shared_busway*, or *cycleway:right=shared_busway*. The OSM wiki for Bicycle (*13*) illustrates numerous variations of lane alignment and orientation with the corresponding tags. The present OSM tagging system does not include a bicycle shoulder, which may be of the same width as a bicycle lane but lack markings or signs. As a consequence, unmarked shoulders are sometimes incorrectly tagged as lanes. Although *shared_lane* and *sharrow* values exist for the cycleway key, we excluded these lane types from analysis since they require the cyclist to share the lane with motorized vehicles.

The code below shows the SQL queries that we composed and used in the PostgreSQL database to extract bicycle trails and designated lanes from the OSM data files.

<u>Bicycle trails:</u>

```
WHERE
((tags->'highway') = 'track' AND ((tags->'bicycle') = 'designated') AND ((tags->'motor_vehicle') = 'no'))
OR ((tags->'highway') = 'path' AND ((tags->'bicycle') = 'yes'))
OR ((tags->'highway') = 'path' AND (((tags->'bicycle') = 'designated') OR (tags->'bicycle') = 'official'))
OR ((tags->'highway') = 'service' AND ((tags->'bicycle') = 'designated') AND ((tags->'motor_vehicle') = 'no'))
OR ((tags->'highway') = 'pedestrian' AND (((tags->'bicycle') = 'yes') OR (tags->'bicycle') = 'official'))
OR ((tags->'highway') = 'footway' AND (((tags->'bicycle') = 'yes') OR (tags->'bicycle') = 'official'))
OR ((tags->'highway') = 'cycleway')
OR ((tags->'highway') = 'bridleway' AND (tags->'bicycle') !='no')
OR ((tags->'cycleway') = 'track')
OR ((tags->'cycleway') = 'opposite_track');
```

Designated lanes:

```
WHERE
((tags->'cycleway') = 'lane')
OR ((tags->'cycleway:left') = 'lane')
OR ((tags->'cycleway:right') = 'lane')
OR ((tags->'cycleway:both') = 'lane')
OR ((tags->'cycleway') = 'opposite_lane')
OR ((tags->'cycleway') = 'shared_busway')
OR ((tags->'cycleway:left') = 'shared_busway')
OR ((tags->'cycleway:right') = 'shared_busway');
```

The first three rows in FIGURE 1 show examples for bicycle trail features that can be identified through the afore-mentioned SQL queries. The right most column lists the tag of each visualized OSM feature. The OSM feature in FIGURE 1a is a bicycle trail that is accessible to cyclists only, whereas the bicycle trail in FIGURE 1b provides also access for pedestrians. FIGURE 1c shows the coding for a cycle track, which we also classify as a bicycle trail since it is separated from the motorized traffic lanes.

FIGURE 1d shows a common combination of motorized traffic lanes with bicycle lanes running in both directions, whereas in FIGURE 1e a one-way cycle lane runs on the opposite direction of the one-way road.

Bicycle trail

a) Amsterdam

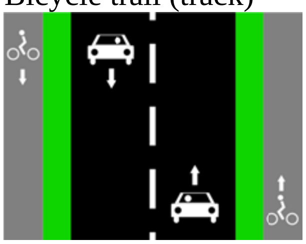
highway=cycleway

Bicycle trail

b) Berlin

highway=cycleway + footway=yes

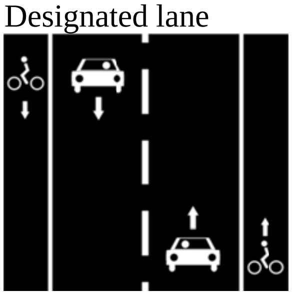
Bicycle trail (track)



c) Berlin

highway=primary + cycleway=track + oneway=yes

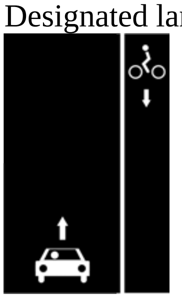
Designated lane



d) San Francisco

highway=tertiary + cycleway=lane

Designated lane



e) Portland

highway=tertiary + oneway=yes + cycleway=opposite_lane

FIGURE 1 Bicycle trails and designated lanes; street images are taken from Google Street View.

STUDY SETUP

Analysis of Cycling Data

The study consists of three analyses for different regions. The first analysis gives an overview of relative completeness of OSM cycling data for selected US regions based on a data dump from 13 June 2012. We use boundaries of the 70 Census 2000 Urbanized Areas with a population larger than 500,000 to clip OSM data and measure the total length of OSM bicycle trails and designated lanes for these areas. The second analysis measures the growth of OSM bicycle data (trails and lanes) in seven metropolitan areas in the United States (Portland, San Francisco, Washington) and Europe (Amsterdam, Berlin, London, Madrid) between the years 2009 and 2012. Third, we assess the completeness of bicycle features in OSM and Google Maps data through data comparison which is performed for a test area in Portland.

Extraction of Bicycle Related Data

The extraction of annual OSM data was conducted using Ubuntu Server and PostgreSQL/PostGIS database software. The OSM full planet files were downloaded from planet.openstreetmap.org for the years 2009 to 2012. The newer and faster data pbf format was only available for the year 2012. All other years were obtained in the traditional osm-xml format. The data for the seven cities was extracted from the planet files by applying the freely available osmosis tool and implementing corresponding bounding box information. After the different databases for each city and year were created in PostgreSQL, the extracted data was imported into the databases using the osmosis tool. Polygons for the seven city boundaries were obtained in shapefile format from different sources, i.e., Centro de Descargas – Centro Nacional de Informacion Gegrafica (Madrid), Centraal Bureau voor de Statistiek (Amsterdam), Ordnance Survey – OS OpenData (London), TomTom (Berlin), U.S. Census Buerau TIGER/Line (all US cities and Urbanized Areas), and imported into the PostgreSQL database. Cyclist layers were then clipped to polygon areas, and the total length of bicycle trails and designated lanes was computed using PostGIS functions. The clipped features were also exported to shapefiles for further analysis in ESRI’s ArcGIS 10.

Google Maps has been providing bicycle related routing features since March 2010 on its Website (*3*). While the data are not available for download in vector format, they can be viewed in a Web browser when activating the Bicycling layer. The layer contains three categories, which are *Trails* (a dedicated bike-only trail), *Dedicated lanes* (a dedicated bike lane along a road), and *Bicycle friendly roads* (roads that are designated as preferred for bicycling, but without dedicated lanes). FIGURE 2 shows the area around the Olympic Stadium in London with the Bicycling layer turned on.

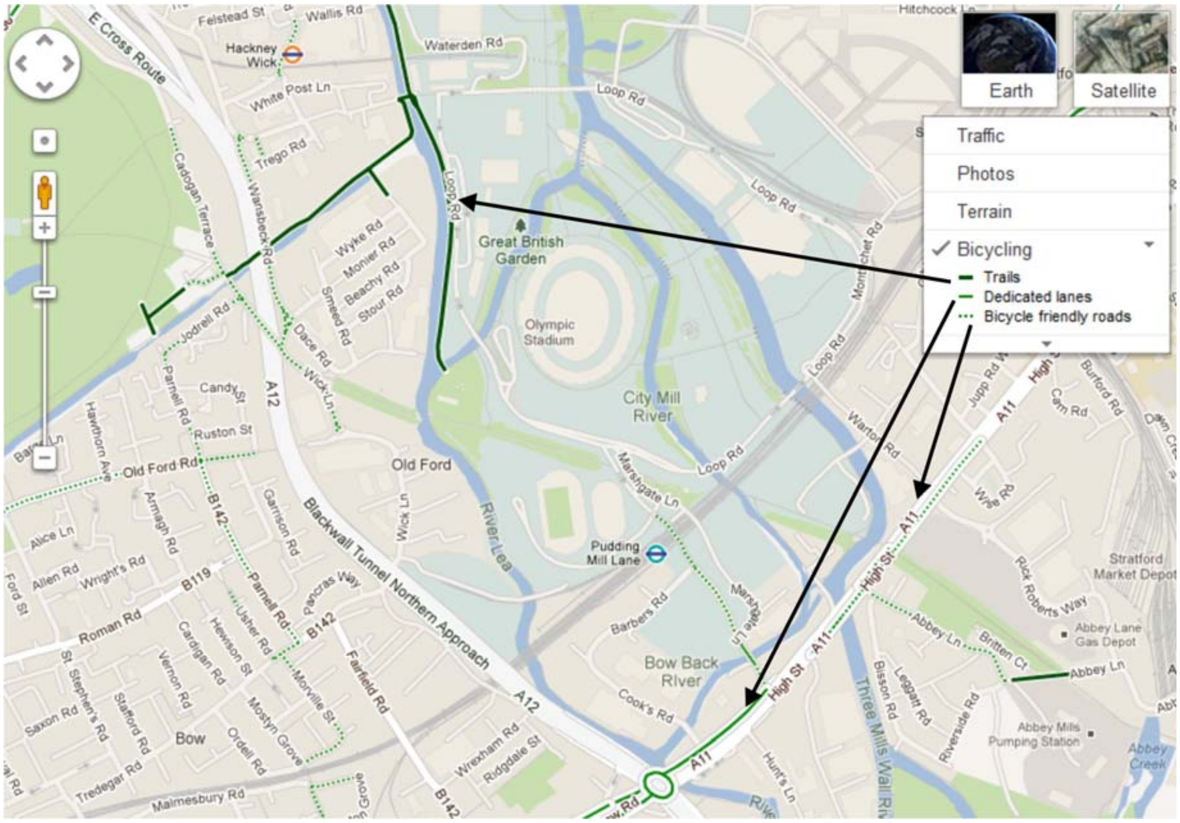


FIGURE 2 Bicycling layer in Google Maps.

While it cannot be assumed that Google Maps provides perfect coverage of cycling features, it is a useful proprietary data reference for comparison with OSM, considering that it is one of the most widely used trip planning Web sites. To assess the usefulness of the Google Maps cycling trails and dedicated lane categories for analysis, we visually compared these layers to reference vector data provided by Broward County Metropolitan Planning Organization (MPO), Florida. A good match was found for bicycle trails. For the dedicated bicycle layer it could be observed that Google generally includes only roads with bicycle lanes that have markings or signage, which are the same criteria that OSM applies to define cycleways. The match of Google's dedicated lanes features with the MPO dataset was generally high. In a few instances roads with three-foot undesignated lanes were included in the Google dedicated lanes category, whereas in a few other cases dedicated bike lanes were missing in Google.

Bicycle trails are mapped as off-road features both in OSM and Google Maps using a center line, which makes a length comparison straight forward. As opposed to this, bicycle lanes can run either on one side or both sides of a road. In OSM the location and direction of lanes along a road can be coded through a namespace for the lane tag, such as *cycleway:right=lane*, which indicates that a lane exists only to the right side of a road. While OSM can provide lane location, the Google Bicycling layer is always mapped through one centerline only without providing further information about the lane location. An exception are roads with medians and split directional lanes for car traffic. Because of this limitation of the visualized Google Maps data, we do not distinguish between left and right lanes, but consider only absence and presence of the center line for bicycle lane analysis in OSM and Google Maps data.

RESULTS

Completeness of OSM Bicycle Features in US Urbanized Areas

TABLE 1 lists for the 70 Urbanized Areas the total length of OSM mapped bicycle trails and designated lanes. The rows are sorted from largest to smallest trail density, computed as total trail km divided by area in km^2. Similarly, the lane density is computed as total trail km divided by area in km^2. Ranks for both quantities are provided as well. The Denver-Aurora area has the highest density of OSM bicycle trails, while Portland shows the highest density of designated lanes in OSM. No correlation was identified between population or size and trail density, and between population or area and designated lane density (Pearson $|r|<0.1$ for all combinations). The correlation between trail and lane densities is moderate (Pearson $r=0.40$, $p<0.001$), meaning that some urbanized areas have high OSM trail densities and lower OSM lane densities at the same time. As an example, the Portland Urbanized Area with a high trail and lane density is visualized in FIGURE 3. The red square in the center of the map indicates the 25 km^2 test area that will be used for manual validation of data completeness for OSM and Google cycling data as described further below.

TABLE 1 OSM Bicycle trails and designated lanes for 70 urbanized areas in the United States.

Urbanized Area	POP 2000	TRAIL Total km	TRAIL Density	TRAIL Rank	LANE Total km	LANE Density	LANE Rank
Denver-Aurora	1982658	1067	2.119	1	50	0.038	23
Portland	1580720	526	1.107	2	1274	1.036	1
Mission Viejo	531816	120	0.872	3	177	0.498	3
Concord	551756	147	0.834	4	23	0.050	18
Chicago	8299353	1764	0.823	5	321	0.058	15
Minneapolis-St. Paul	2385465	757	0.805	6	109	0.045	21
San Francisco-Oakland	2985722	339	0.798	7	325	0.295	6
Washington	3936201	741	0.626	8	140	0.046	20
San Jose	1537148	148	0.570	9	566	0.840	2
Baltimore	2076163	372	0.540	10	22	0.012	34
Sacramento	1392438	198	0.537	11	321	0.335	4
Tulsa	557007	122	0.464	12	8	0.011	35
Dayton	701872	140	0.433	13	4	0.004	48
Phoenix-Mesa	2904968	342	0.427	14	228	0.110	13
Kansas City	1358784	241	0.411	15	23	0.015	31
Raleigh	539799	130	0.403	16	0	0.000	56
Los Angeles-LB..-Sta. Ana	1178447	640	0.380	17	1110	0.255	7
Seattle	2708131	363	0.376	18	141	0.056	16
Allentown-Bethlehem	572732	105	0.363	19	8	0.011	37
Buffalo	975126	125	0.341	20	25	0.027	27
Orlando	1155470	150	0.309	21	302	0.241	8
San Antonio	1325146	121	0.296	22	1	0.001	54
New Orleans	1008509	55	0.275	23	41	0.079	14
Albany	556316	76	0.267	24	5	0.006	43
Salt Lake City	886572	60	0.258	25	101	0.169	12
Austin	899689	79	0.247	26	9	0.011	38
Boston	4014865	425	0.240	27	57	0.012	33
Milwaukee	1306729	118	0.237	28	20	0.016	30
Columbus	1129836	91	0.228	29	6	0.006	44
Rochester	692839	63	0.211	30	0	0.000	56
Tampa-St. Petersburg	2049318	167	0.204	31	467	0.220	9
Oklahoma City	745352	65	0.198	32	0	0.000	56
Fresno	554239	27	0.195	33	7	0.020	29
Las Vegas	1312198	55	0.192	34	0	0.000	56
Hartford	850234	88	0.186	35	2	0.001	53
Providence	1166118	89	0.174	36	10	0.007	40
Springfield	570748	52	0.166	37	0	0.000	56
St. Louis	2075336	136	0.163	38	5	0.002	50
Dallas-Fort Worth-Arl.	4140851	224	0.159	39	0	0.000	56
Detroit	3900539	199	0.154	40	14	0.004	47
San Diego	2669584	120	0.151	41	75	0.037	24
Pittsburgh	1746379	128	0.150	42	14	0.007	42
New York-Newark	1734004	472	0.141	43	233	0.027	26
Philadelphia	5142385	247	0.137	44	185	0.040	22
New Haven	528784	40	0.134	45	8	0.011	36
Grand Rapids	538761	33	0.127	46	0	0.000	56
Toledo	502146	26	0.127	47	0	0.000	56
Akron	568432	39	0.125	48	0	0.000	56
Cincinnati	1500552	82	0.122	49	5	0.003	49
Riverside-San Bernardino	1504093	52	0.118	50	56	0.049	19
Houston	3819632	152	0.117	51	17	0.005	46
Virginia Beach	1391122	62	0.113	52	7	0.005	45
Albuquerque	597398	21	0.096	53	17	0.028	25
Miami	4901994	102	0.090	54	504	0.171	11
Indianapolis	1214928	46	0.083	55	34	0.023	28
Sarasota-Bradenton	555681	21	0.076	56	218	0.313	5

Cleveland	1785038	47	0.072	57	12	0.007	41
Atlanta	3493117	125	0.063	58	40	0.008	39
Tucson	719452	18	0.062	59	39	0.052	17
Jacksonville	880960	23	0.055	60	221	0.207	10
Louisville	862163	15	0.038	61	0	0.000	56
El Paso	673865	8	0.034	62	1	0.002	51
Memphis	970689	10	0.025	63	15	0.015	32
Omaha	625805	6	0.024	64	0	0.000	56
Bridgeport-Stamford	884229	7	0.015	65	0	0.000	56
Charlotte	753867	6	0.014	66	0	0.000	55
Nashville	747512	5	0.011	67	0	0.000	56
Richmond	817525	2	0.005	68	0	0.000	56
Birmingham	661177	1	0.003	69	2	0.002	52
McAllen	520667	0	0.000	70	0	0.000	56

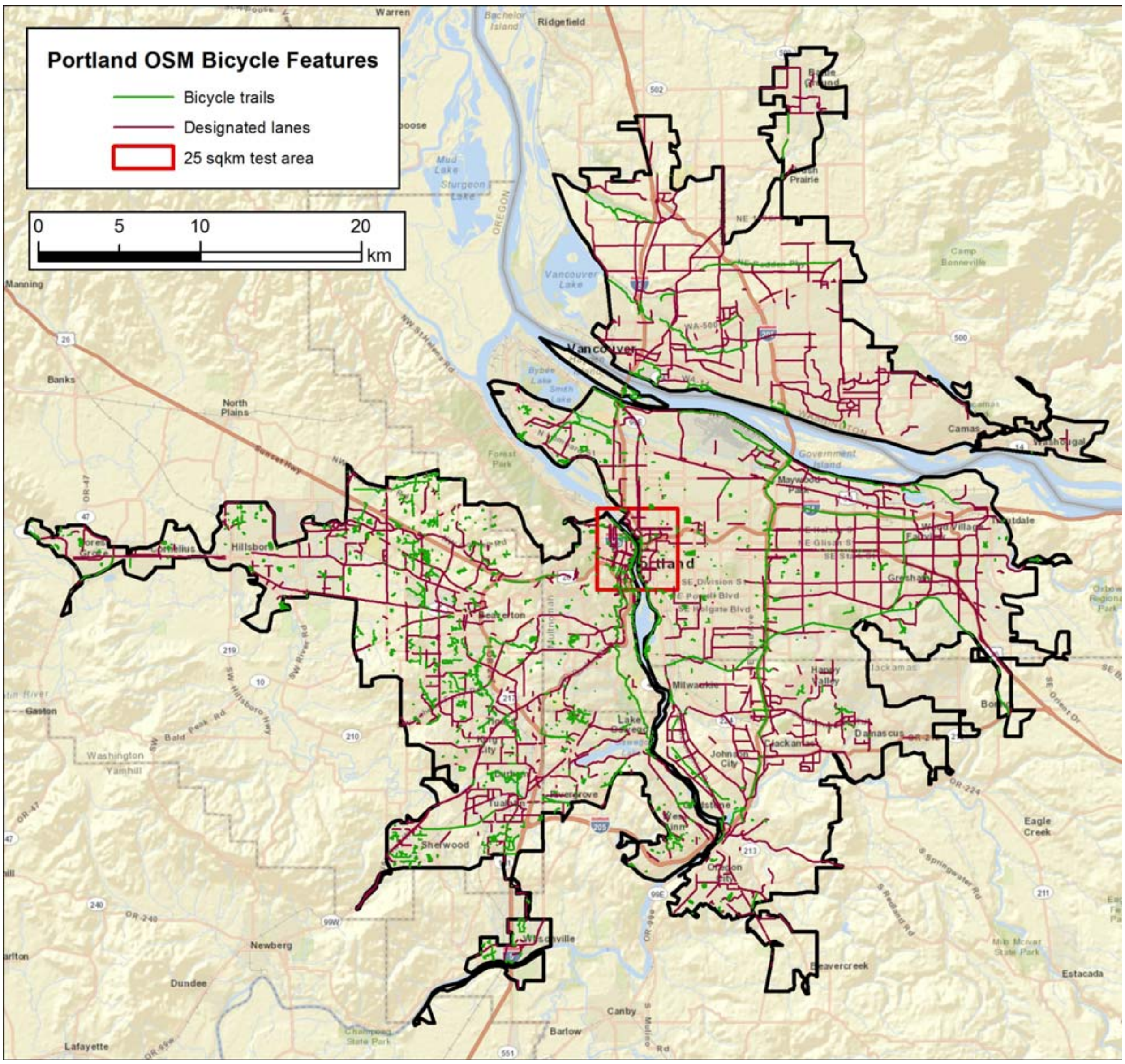


FIGURE 3 OSM bicycle trails and designated lanes in the Portland Urbanized Area.

Growth of OSM Bicycle Feature Data in Selected US and European Cities

FIGURE 4 shows the growth of total length for bicycle trails and designated lanes in the OSM data base for seven US and European metropolitan areas between 2009 and 2012. The diagrams to the left express total kilometers for each area, those to the right show densities. The increasing values for the lines over time indicate that the OSM mapping community is active, however, to a different degree in different cities. Amsterdam and Berlin stand out with high contribution rates for bicycle trails over the years (FIGURE 4a, b), whereas the three US cities, particularly Portland, show a higher absolute increase in density of designated lanes (FIGURE 4c, d). In a few cases some cities show a decline in coverage in consecutive years, for example, designated lanes for Washington, DC between 2011 and 2012. This decline is caused by correcting tags from prior years and removal of incorrectly tagged designated lanes.

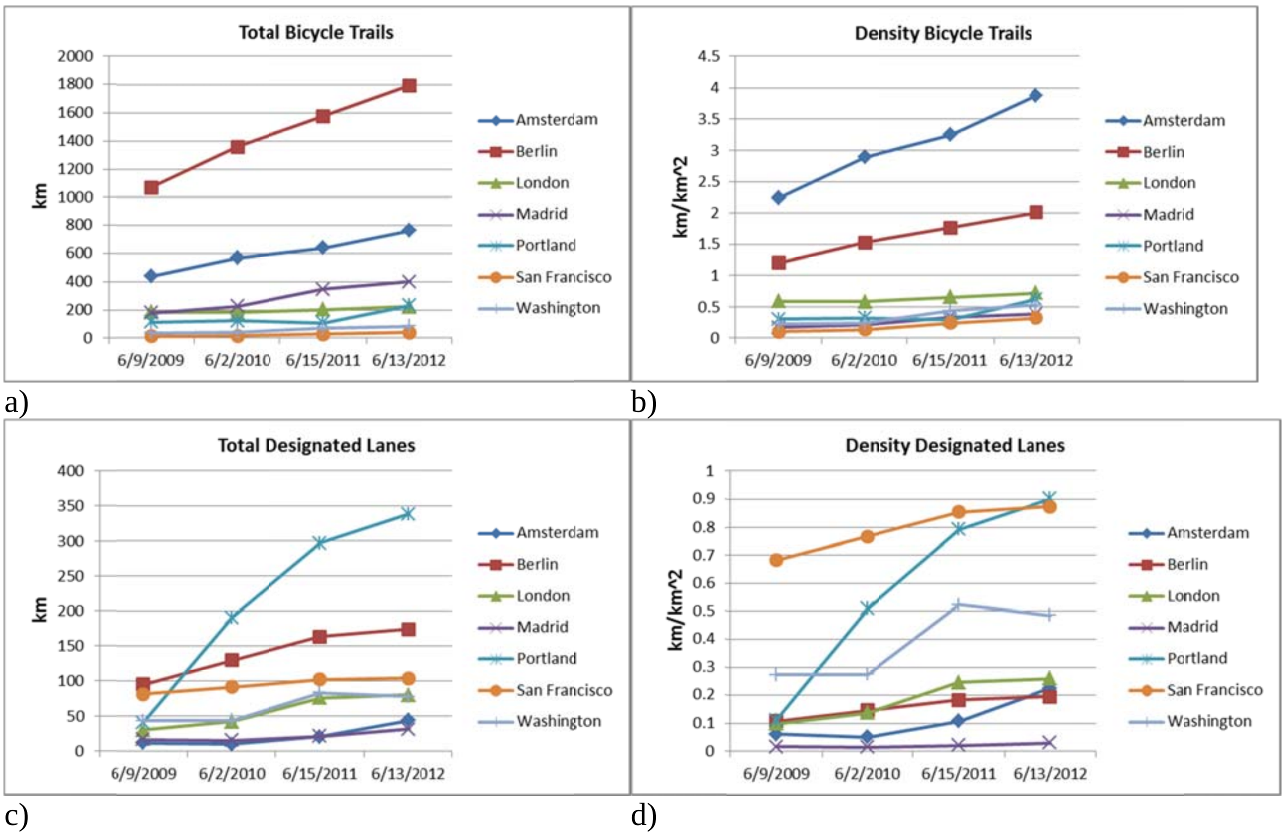


FIGURE 4 Growth of OSM cycling data between 2009 and 2012.

Bicycle trail densities are larger for Amsterdam and Berlin compared to London, Madrid, and the three US cities (FIGURE 4a, b). One possible contributing factor to this difference is that Amsterdam and Berlin oftentimes provide separate cycling tracks along arterial roads, e.g. separated through grass and tree islands (compare FIGURE 1c), which are therefore classified as bicycle trails in our analysis. This roadway design is not as prominent in the other cities. The higher rate of bicycle trails could also explain the lower rate of designated lanes found in these two cities, at least when compared to the US cities (FIGURE 4c, d). One factor for low densities (trail and lane) for Madrid could be the large reference area (it is the largest with 1046 km^2), which includes an unpopulated mountain region in the north.

Comparison of OSM and Google Bicycling Data Completeness

This section describes a manual validation approach which was used to quantify the completeness of OSM and Google cycling features, based on a 25 km^2 test area around the Portland Central Business District (indicated in FIGURE 3). The geometric correctness of OSM and Google bicycle features (trails and lanes) can be quantified through two types of errors with reference to ground truth data, i.e., error of omission and error of commission. An error of omission occurs if the bicycle feature is missing in the digital dataset, while it exists in the real world. The error can be caused by a missing geometry or by an incorrect tagging. Missing geometry means that the feature geometry is not present in the dataset at all, not even for a different highway type. As opposed to this, incorrect tagging means that the correct feature geometry is present in the dataset, but that an incorrect tag prevents the feature to be extracted as the corresponding cycling feature within the query.

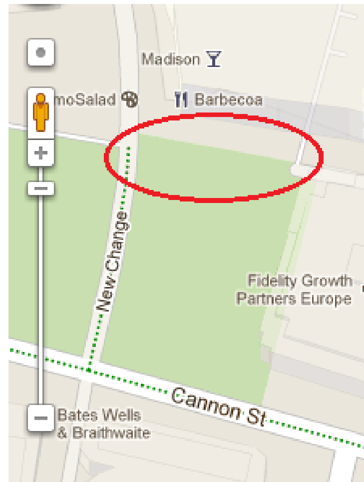
An error of commission occurs if a cycling feature is present in the dataset but if it does not exist in the real world. Also, this error can be caused by a missing geometry or incorrect tagging. Incorrect geometry means that a cycling feature is mapped in the dataset, but that no linear feature can be found in the real environment for that location. Incorrect tagging makes a feature that is not related to cycling, e.g. a footpath for pedestrians only, to be coded and mapped as a cycling feature, e.g. a bicycle trail.

FIGURE 5 provides for the London road network some examples of these errors. FIGURE 5a shows in Google Street View a footpath that also can be traversed by bicycle, as shown by the bicycle path sign on the pillar to the right. FIGURE 5b maps the situation for the OSM data. The highlighted path corresponds to the one shown in the Google Street View image. Its tags correctly state that bicycles are allowed. As opposed to this, no geometry is shown for this alley in Google maps (FIGURE 5c). Therefore this is an error omission with missing geometry in the Google road data.

The second situation is a street without bicycle facilities, as shown in FIGURE 5d. While the Google Maps bicycle layer is correct in not showing a bicycle lane for this street (FIGURE 5f), the OSM tags of this road suggest a bicycle lane (FIGURE 5e). This is therefore an error of commission for the OSM dataset caused by incorrect tagging of that road.

Error of omission:

highway=footway + bicycle=yes

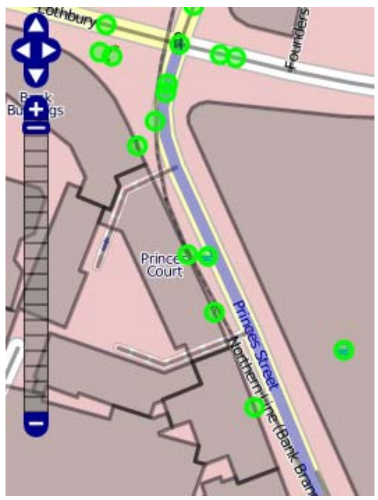


a) Google Street View

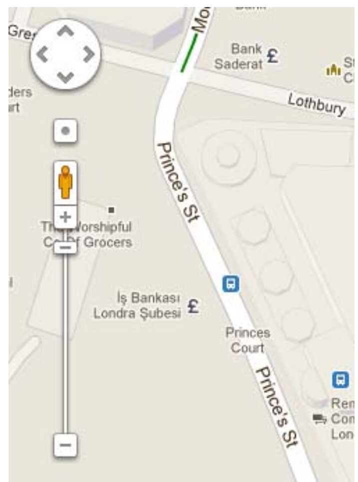
b) OSM

c) Google Maps: Error of omission (missing geometry)

Error of commission:



highway=tertiary + cycleway=lane



d) Google Street View

e) OSM: Error of commission (incorrect tag)

f) Google Maps

FIGURE 5 Error of omission and error of commission.

For the assessment of data completeness a reference data set is required that represents ground truth. This would be primarily necessary to detect all missing cycling elements in the OSM and Google datasets. As opposed to this, it is less problematic to detect an error of commission, i.e., identify a cycling feature in the data set that is not present in the real world, using alternative data sources. We use the Google Maps back ground image and Google Street View to identify errors of commission. For errors of omission, although not a perfect solution, we use a combination of OSM and Google Maps bicycle features as auxiliary ground truth dataset. More specifically, we go through each individual mapped cycling feature in OSM and Google Maps (trail and designated/dedicated lane) and check whether it exists in the real world through satellite imagery and Google Street View. If so, we consider the cycling feature to be present at the mapped location and it is added to the ground truth map as reference. This approach provides a good measure of relative completeness of each dataset with respect to the combination of the two datasets.

For the validation approach four rounds of visual checks were conducted for errors of omission and commission of bicycle features using aerial imagery and Google Street View. Following four layers were checked: (1) OSM bicycle trails, (2) Google bicycle trails, (3) OSM designated lanes, (4) Google dedicated lanes. TABLE 2 summarizes the analysis results for the Portland test area. The upper half refers to bicycle trail assessment, the lower one to the assessment of cycling lanes. The first “Total” value in each half describes to the total length of auxiliary ground truth trails and lanes, respectively. Ground truth data consists of correctly mapped features of a data source and features that were identified as omitted (through check with another data source). About 26 km of trail features were found in the test area, out of which 86.6% were correctly retrieved in OSM and 75.0% in Google Maps. Thus OSM misses in this sample fewer cycling trail features than Google Maps. As the error breakdown shows in the same line, omitted features can be mostly attributed to incorrect tags, especially in the OSM case. In Google, separated tracks are sometimes coded as dedicated lanes instead of trails, which gives the relatively high error of omission for bicycle trails. For designated/dedicated lanes (see lower half) the length of auxiliary ground truth lane features totals 59 km. Compared to trails, a higher percentage of lanes is correctly captured (around 97% for both datasets), and fewer features are omitted.

The rates for error of commission show the total length of features that were in the dataset classified as trail or lane, respectively, but could not be identified on the ground. The percentage values for this error relate to the total distance of auxiliary ground truth features provided above. Most errors of commission were caused by incorrect tags. While rates for error of commission are low and in the range of 3-4% for all Google data and OSM designated lanes, they are higher for OSM trails with 14.5%.

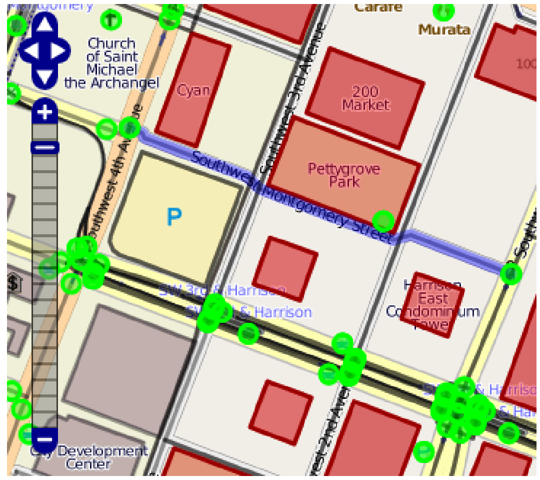
TABLE 2 Assessment of completeness for OSM and Google data in the Portland test area.

TRAILS	**OSM**				**Google**			
		Error of omission				*Error of omission*		
	Correct	Tag	Geometry	**Total**	Correct	Tag	Geometry	**Total**
[m]	22981	*3097*	*475*	**25553**	19905	*3846*	*2802*	**25553**
[%]	86.6	*11.7*	*1.8*	**100**	75.0	*14.5*	*10.6*	**100**
		Error of commission				*Error of commission*		
		Tag	Geometry	**Total**		Tag	Geometry	**Total**
[m]		*3678*	*171*	**3849**		*875*	*154*	**1029**
[%]		*13.9*	*0.6*	**14.5**		*3.3*	*0.6*	**3.9**

LANES	**OSM**				**Google**			
		Error of omission				*Error of omission*		
	Correct	Tag	Geometry	**Total**	Correct	Tag	Geometry	**Total**
[m]	57274	*1897*	*0*	**59171**	57536	*1538*	*97.4*	**59171**
[%]	96.8	*3.2*	*0*	**100**	97.2	*2.6*	*0.2*	**100**
		Error of commission				*Error of commission*		
		Tag	Geometry	**Total**		Tag	Geometry	**Total**
[m]		*1964*	*0*	**1964**		*2099*	*212*	**2311**
[%]		*3.3*	*0*	**3.3**		*3.5*	*0.4*	**3.9**

While there are different causes for errors of commission with the OSM trail dataset, two typical cases can be observed. The first one is to incorrectly tag a footpath that is accessible to pedestrians only (e.g. a narrow sidewalk) with bicycle access. An example is provided in FIGURE 6 (upper row). In FIGURE 6a the footpath has stairs and can therefore not be accessed by bicycle. However, the OSM tagging suggests that the path is accessible to cyclists (FIGURE 6b). The second case is a bicycle feature that is misclassified. FIGURE 6c shows a road with an on-road bicycle lane. However the road is tagged with *cycleway=track* in OSM (FIGURE 6d), which is to be used for spatially separated bike lanes only. This provides therefore a false positive for a cycling trail.

Incorrect tagging of mode



highway=pedestrian +
bicycle=yes

a) Google Street View b) OSM error of commission

Incorrect tagging of facility

highway=secondary +
cycleway=track

c) Google Street View d) OSM error of commission

FIGURE 6 Errors of commission in the OSM dataset.

Implications for Practitioners

Given the limited choice of resources for road network datasets in vector format that provide cycling information, OSM proves to be a relatively accurate, free, and easily accessible data alternative for certain types of analyses. Customization of SQL queries allows to further filter road and feature types.

The results of this quality assessment study reveal that, at least for the Portland test area, designated OSM bicycle lanes largely overlap with Google Maps dedicated lanes, and only a small percentage of designated lanes is incorrect. For off-road trails OSM is still relatively powerful in retrieving relevant features. However about 14% of trail features are falsely classified as such. In this regard, the Google data outperform OSM (fewer false positives), and it is recommended to visually compare OSM extracted trails with those from Google Maps to reduce the rate of false positives. OSM provides also the opportunity to integrate local cycling data into the OSM database using various data editors, such as JOSM (http://josm.openstreetmap.de/). This way, local cycling data which are missing in OSM could be added to OSM and made accessible to the public. The presented test was performed in an area that ranks high in OSM data coverage. It is therefore recommended to compare OSM completeness with Google Maps bicycling layers in particular when working in other areas for which OSM provides less complete cycling data.

SUMMARY AND FUTURE WORK

This paper analyzed the development of cycling related feature data in OpenStreetMap between 2009 and 2012, indicating that the OSM mapping community is active and that the amount of mapped cycling data is likely to grow even more in the near future. Comparison between selected US and European cities showed that Amsterdam and Berlin have a large network of bicycle trails, while the selected US cities have a higher density of designated lanes. A review of 70 Urbanized Areas in the US demonstrates the large range of trail and lane densities within this set. The results can provide some hints about the bicycle friendliness in these different areas, given the assumption that the OSM mapping community is similarly active in these regions.

This study was focusing on OSM bicycle trails and designated lanes. While these features are important for various transportation planning and analysis tasks, additional road attributes, such as traffic volume, and facility features, such as undesignated lanes, would be necessary for the computation of the bicycle Level of Service (BLOS), many of these attributes and facility features are currently not present in the OSM coding scheme (*5*). For future work it therefore is necessary to observe and analyze the future development of additional bicycle related attributes and features in the OSM coding schemes, which can usually be first found in OSM tagging proposals suggested by OSM data contributors. Another aspect of future work is to customize existing OSM data import tools so that local data can be easier integrated into the OSM data base and made accessible to the public.

References

(*1*) Barnes, G. and Krizek, K. Estimating Bicycling Demand. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. *1939*, 2005, pp. 45-51.

(*2*) Dill, J. and Carr, T. Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters Will Use Them – Another Look. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. *1828*, 2003, pp. 116-123.

(*3*) Guymon, S. Google Official Blog: Biking directions added to Google Maps. http://googleblog.blogspot.com/2010/03/biking-directions-added-to-google-maps.html. Accessed 11/7/2012

(*4*) Haklay, M. How good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B, Planning and Design*, Vol. *37*, 4, 2010, pp. 682 – 703.

(*5*) Hillsman, E. L. and Barbeau, S. J. *Enabling Cost-Effective Multimodal Trip Planners through Open Transit Data* Tallahassee, FL. Available from *http://ntl.bts.gov/lib/38000/38500/38506/FDOT_BDK85_977-20_rpt.pdf*, 2011.

(*6*) Hochmair, H. H. *GIS-based Identification of Need for Bicycle Level of Service Improvement in Street Networks*. 12th AGILE International Conference on Geographic Information Science, Hannover, Germany, 2009.

(*7*) Hochmair, H. H. and Fu, J. *Web Based Bicycle Trip Planning for Broward County, Florida*. ESRI User Conference, San Diego, CA. (CD-ROM), 2009.

(*8*) Krizek, K. J. and Johnson, P. J. Proximity to Trails and Retail: Effects on Urban Cycling and Walking. *Journal of the American Planning Association*, Vol. *72*, 1, 2006, pp. 33-42.

(*9*) Landis, B. W. Bicycle System Performance Measures. *ITE Journal*, Vol. *66*, February 1996, 1996, pp. 18-26.

(*10*) Mooney, P. and Corcoran, P. Characteristics of Heavily Edited Objects in OpenStreetMap. *Future Internet*, Vol. *4*, 2012, pp. 285-305.

(*11*) Mooney, P. and Corcoran, P. *How social is OpenStreetMap?* AGILE 2012, Avignon, 2012.

(*12*) Neis, P., Zielstra, D., and Zipf, A. The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. *Future Internet*, Vol. *4*, 1, 2011, pp. 1-21.

(*13*) OSM OSM wiki for Bicycle. http://wiki.openstreetmap.org/wiki/Bicycle#Cycle_tracks. Accessed 11/7/2012

(*14*) OSM OSM wiki for Map Features. http://wiki.openstreetmap.org/wiki/Map_Features. Accessed 11/7/2012

(*15*) Ramm, F. and Stark, H.-J. Crowdsourcing geodata. *Geomatik Schweiz, Géomatique Suisse*, Vol. *2008*, 6, 2008, pp. 315-318.

(*16*) Ramm, F., Topf, J., and Chilton, S. *OpenStreetMap: Using and Enhancing the Free Map of the World*, UIT Cambridge, Cambridge, 2010.

(*17*) Rehrl, K., Gröchenig, S., Hochmair, H. H., Leitinger, S., Steinmann, R., and Wagner, A. A Conceptual Model for Analyzing Contribution Patterns in the Context of VGI. In J. M. Krisp (Ed.), *Progress in Location Based Services*, *Lecture Notes in Geoinformation and Cartography*, Berlin, Springer, 2013.

(*18*) Su, J. G., Winters, M., Nunes, M., and Brauer, M. Designing a route planner to facilitate and promote cycling in Metro Vancouver, Canada. *Transportation Research Part A: Policy and Practice*, Vol. *44*, 2010, pp. 495–505.

(*19*) Zielstra, D. and Hochmair, H. H. A Comparative Study of Pedestrian Accessibility to Transit Stations Using Free and Proprietary Network Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. *2217*, 2011, pp. 145-152.

(*20*) Zielstra, D. and Hochmair, H. H. Comparing Shortest Paths Lengths of Free and Proprietary Data for Effective Pedestrian Routing in Street Networks. *Transportation Research Record: Journal of the Transportation Research Board*, in press.