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

Hartwig H. Hochmair¹*, Dennis Zielstra¹, Pascal Neis²

¹University of Florida, Geomatics Program, Fort Lauderdale Research and Education Center
²University of Heidelberg, Geography Department, Geoinformatics Research Group

*Author to whom correspondence should be addressed:

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

Short title: Completeness of Bicycle Features in OpenStreetMap

Key words: OpenStreetMap, Volunteered Geographic Information, cycling, data completeness, bicycle infrastructure
ABSTRACT
This paper assesses the completeness of bicycle trail and on-street lane features in OpenStreetMap (OSM). Comparing OSM cycling features with reference data from local planning agencies for selected US Urbanized Areas shows that OSM bicycle trails tend to be more completely mapped than bicycle lanes. Manual evaluation of mapped cycling features in OSM and Google Maps for selected test areas within the Central Business Districts of Portland (OR) and Miami (FL) through comparison with governmental datasets, satellite imagery, and Google Street View, shows that the Bicycle layer in Google Maps can help to identify some missing or erroneously mapped OSM cycling links. However, also Google Maps was found to have some gaps in its data layers, suggesting that consultation of current trail and lane data from local planning authorities as an additional data source should be considered for bicycle related planning projects if available.
1. INTRODUCTION
Digital data repositories of bicycling related information in a transportation network, such as on-street bicycle lanes, off-street trails, lane width, or traffic volume, provide an important basis for a variety of bicycle transportation planning and analysis tasks, including latent demand analysis (Barnes and Krizek, 2005; Hochmair, 2009; Landis, 1996), bicycle trip planning applications (Hochmair and Fu, 2009; Su et al., 2010), travel behavior analysis (Dill and Carr, 2003; Krizek and Johnson, 2006) and estimation of Bicycle Level of Service (BLOS) (Landis et al., 1997) or Bicycle Compatibility Index (BCI) measures (Harkey et al., 1998) for road segments. There exist a variety of public and proprietary street data sets that can be used as basis for the analysis of motorized traffic and underlying network structures. 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 for Germany from 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 includes network data for cyclists, such as bicycle trails or on-street lanes. Transportation planning tasks at the local level that involve non-motorized traffic, including cycling, are therefore usually based on data collected by local stakeholders, e.g. county and city planning departments, and in consequence fragmented and not easily accessible to the public. OpenStreetMap (OSM) (www.openstreetmap.org), which is an open, freely available repository of geographic data, provides a feasible alternative to scattered, locally stored cycling data. More specifically, OSM uses widely accepted, since community developed, coding conventions for cycling related information in a transportation network and hosts numerous transportation feature layers in a
publicly accessible and editable database. The OSM data volume is rapidly growing, and new
data tagging proposals from OSM data contributors towards a more refined mapping of cycling
relevant facilities and road attributes are underway, making OSM a promising data repository for
the cyclist community.

The objective of this paper is to assess the completeness of bicycle related network data
in OSM. The geographic focus lies on larger US cities, for which a recent study compiled the
total length of bicycle trails and lanes in 2008 (Buehler and Pucher, 2012). This can be used as a
reference dataset for comparison with OSM mapped trails and lanes. While road network
variables provided by the OSM tagging scheme are not (yet) detailed enough to compute Bicycle
Level of Service measures (Hillsman and Barbeau, 2011), bicycle trails and designated on-street
lanes are already coded, which facilitates a basic level of analysis and mapping of the cycling
network with OSM data (e.g. http://www.opencyclemap.org). The results of this research will
contribute to a better understanding of the data quality of OSM cycling data for analysis and
mapping tasks in the U.S.

2. OPENSTREETMAP
OpenStreetMap is a collaborative mapping project with data contributed by voluntary users. 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 OSM 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 market.weogeo.com and geofabrik.de provide OSM data
download for pre-selected administrative areas, such as continent, country, or state, in common
GIS formats including shapefiles. OSM hosts spatial data from a large selection of themes,
including roads, transit lines, tourist sites, or land use layers. One direction of OSM research is concerned with user contribution patterns, i.e., what kind of data different users contribute and how frequently (Mooney and Corcoran, 2012b; Neis and Zipf, 2012; Ramm and Stark, 2008). Another research direction focuses on the assessment of data quality. This is particularly relevant for OSM because its data are not administered by a regulatory instance but contributed by non-professional individuals. Previous studies analyzed among others data completeness (Neis et al., 2012; Zielstra and Hochmair, 2011a, 2012), positional accuracy (Haklay, 2010), attribute accuracy (Mooney and Corcoran, 2012a; Neis et al., 2012), temporal development (Corcoran et al., 2013; Rehrl et al., 2013; Zielstra et al., 2013), or several of these aspects together (Girres and Touya, 2010). Among the numerous findings that emerge from the analysis of OSM data completeness it was shown that roads in urban areas are more completely mapped than in rural areas (Haklay, 2010; Neis et al., 2012), that OSM data completeness varies between world regions (Neis et al., 2013), and that OSM performs well in the mapping of pedestrian paths when compared to proprietary and governmental data providers, both for the US (Zielstra et al., 2013) and Europe (Zielstra and Hochmair, 2011a; Zielstra and Zipf, 2010).

Since completeness of OSM street data has so far been primarily studied for motorized and pedestrian navigation, the objective of this study is to analyze data completeness of mapped bicycle features in OSM. This is of interest to transportation planners, given that bicycle commuter rates in the U.S., especially in cities which have built cycling infrastructure and implemented programs to promote cycling and improve cycling safety, have been increasing over the past years (Pucher and Buehler, 2011).

2.1 Tagging of OpenStreetMap Features
In OSM coding, elements (nodes, ways, and relations) are described through tags. OSM employs an open tagging system, which does not enforce standard tags or specifications (Brando and Bucher, 2010). The open tagging system makes it necessary to acknowledge that the same type of feature, e.g. a bicycle trail, can be tagged differently by different users. Therefore OSM provides guidance for best practices and typical coding schemes. Each tag consists of a key and a value and is written as Key=value in the OSM documentation. 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, where 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 (OSM, 2013c). Comprehensive manuals for using and contributing OSM data are the books by Ramm et al. (2010) and Bennett (2010). A brief introduction to relevant tags for multi-modal transportation, including cycling, is provided by Hillsman and Barbeau (2011).

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 locations such as an industrial estate, camp site, business park, or car park.

2.2 Bicycle Trails and Designated Lanes
With respect to bicycle infrastructure our analysis distinguishes between bicycle trails and designated lanes. Google Maps provides the same distinction in its Bicycling layer. 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. To be consistent with the feature classification in Google Maps, we include in the bicycle trails category also bicycle paths along roads that are physically separated from car traffic, which is referred to as a cycle track in OSM.

There are several options to tag a bicycle trail in OSM. Oftentimes, a bicycle trail is tagged as `highway=cycleway`, which indicates that the way is mainly or exclusively used 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 for bicycle use) 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. 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 one direction, the cycleway:right=lane or cycleway:left=lane tags can be used. The cycleway=opposite_lane tag maps a lane where cyclists are allowed to travel in the direction opposite of other traffic. There are also proposed (but not commonly agreed upon) key-value pairs for cycle lanes shared with roads that have a shared lane marking (sharrow), e.g. cycleway=shared_lane, or for cycle lanes shared with bus and taxi lanes, such as cycleway=shared_busway (OSM, 2013b). While there are tagging schemes available to record wide shoulders in OSM, no consensus has so far been reached among OSM users whether wide shoulders should be mapped at all, since they do not meet current standards for bicycle lanes (Hillsman and Barbeau, 2011). Figure 1 shows the SQL queries that were run in the PostgreSQL database to extract bicycle trails and designated lanes from the OSM data files.

The first three rows in Figure 2 show examples of bicycle trail features that can be extracted through the aforementioned SQL queries. The right most column lists the tags of the visualized OSM feature. The OSM feature in Figure 2a is a bicycle trail that is accessible to cyclists only, whereas the bicycle trail in Figure 2b provides also access for pedestrians. Figure 2c shows the coding for a cycle track, which is visible to the right of the road.

Figure 2d shows a common combination of motorized traffic lanes with bicycle lanes running in both directions, whereas in Figure 2e a one-way cycle lane runs on the opposite direction of the remaining traffic on a one-way street.

3. STUDY SETUP
3.1 Study Area and Overview of Analysis
The study consists of two analyses for different regions. The first analysis gives an overview of relative completeness of OSM cycling data based on a data dump from March 13 2013. More specifically, the analysis measures the total length of OSM bicycle trails and designated lanes inside the boundaries of the 78 US Census 2010 Urbanized Areas with a population larger than 500,000. As part of this analysis we compare also the density of trail and lane features with 2008 reference data (Buehler and Pucher, 2012).

The second analysis assesses for selected test areas in Portland and Miami the completeness of bicycle features in OSM and Google Maps relative to local governmental reference data through manual data comparison. Figure 3 and Figure 4 show the boundaries of the Portland and Miami Urbanized Areas together with OSM trails and designated lanes. The red square in each map indicates the 25 km² test area for the second analysis task.

3.2 Extraction of Bicycle Related Data
The extraction of OSM data was conducted using Ubuntu Server and PostgreSQL/PostGIS database software. The OSM full planet file was downloaded from planet.openstreetmap.org for 2013 in XML format and transformed into a faster binary file format (PBF) with the help of a self-developed Java tool. The data for the 78 Urbanized Areas was extracted from the planet file by applying the freely available OSMOSIS tool (http://wiki.openstreetmap.org/wiki/Osmosis) and implementing corresponding bounding box information. After the database was created in PostgreSQL, the extracted data was imported into the database using the OSMOSIS tool. Bicycle trail and lane layers were then clipped to polygon areas, and the total feature lengths for both layers were computed for each boundary polygon using PostGIS functions and designated tools written in Java. The clipped features were also exported to shapefiles for further analysis in ESRI’s ArcGIS 10.
For the second task we utilized in addition to previously mentioned resources cycling features in Google Maps which have been mapped in the Google Bicycling layer since March 2010 (Guymon, 2010). While these features are not available for download, they can be viewed in a Web browser when activating the Google Bicycling layer. This layer contains three categories, which are *Trails* (a dedicated bike-only or multi-purpose trail), *Dedicated lanes* (a dedicated bicycle lane along a road), and *Bicycle friendly roads* (roads that are designated as preferred for bicycling, but without dedicated lanes).

Bicycle lanes can either run on one side of a road or on both. 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`. As opposed to this, dedicated lanes in the Google Bicycling layer are always mapped through a centerline, whether a lane exists only on one side of the road or on both sides. Exceptions are roads with medians and split directional lanes for car traffic. Due to Google’s mapping scheme we do not count left and right lanes in OSM separately for the computation of total length but count the lane length only once. As a reference dataset for the manual quality assessment in the second task we use GIS layers of bicycle trails and lanes that were obtained from local administrative agencies in Portland and Miami. For Portland the citywide network of bicycle facility types (dated March 2013) was downloaded from the City of Portland Web site (City of Portland, 2013) in shapefile format, which contains bicycle boulevards, bike lanes, multi-use trails and signed connections. For Miami a recent layer file of bicycle trails and lanes was obtained from the Miami-Dade Metropolitan Planning Organization (MPO).

4. RESULTS

4.1 Completeness of OSM Bicycle Features in US Urbanized Areas
Table 1 lists for the first and last 10 of the 78 major Urbanized Areas in the US the total lengths of OSM mapped bicycle trails and designated lanes, where the rows are sorted from largest to smallest trail density, which is computed as total trail km divided by area in km\(^2\). Similarly, the lane density is computed as total lane km divided by area in km\(^2\). Table 1 provides also rank numbers for trail and lane densities, where a smaller rank number indicates a higher density. The Omaha (NE) area has the highest density of OSM bicycle trails, and Portland (OR) shows the highest density of designated lanes in OSM. No significant correlation was identified between population or area and trail density, and between population or area and designated lane density (Pearson \(r<0.1\)). The correlation between log normalized trail and lane densities is moderate (Pearson \(r=0.4, p<0.003\)), indicating that some Urbanized Areas with high OSM trail densities have low OSM lane densities and the other way round.

Densities of mapped trails and dedicated lanes in OSM are a result of (1) the existing bicycle infrastructure in an Urbanized Area, and (2) the completeness of the OSM dataset. Thus, a low density value in Table 1 does not necessarily mean a poor level of OSM mapping completeness. To separate the two effects, we obtained bicycle trail and lane supply values from an alternative data source. While we were not able to obtain current bicycle facility data for all analyzed 78 Urbanized Areas, Buehler and Pucher (2012) provide total lengths of bicycle paths and lanes for the 90 largest US cities in 2008, many of which overlap with the 78 Urbanized Areas. The data in (Buehler and Pucher, 2012) were obtained through contacting bike planners, transportation officials, and bicycling experts in the 100 largest cities of the US. Trail and lane densities computed from this alternative data source can be found in the “Ref. Dens.” columns in Table 1. These data and the OSM density values are not entirely comparable due to the different years of recording (2008 vs. 2013) and the different geographic entities (Urbanized Area vs.
The 2008 reference dataset reveals a non-significant bivariate correlation (Pearson $r=0.2$) between log normalized bicycle path and lane supply for the 90 largest US cities, which is lower than the $r=0.4$ value found for OSM data.

Figure 5 plots the OSM density data against that of the 2008 reference density data for trails (Figure 5a) and lanes (Figure 5b) for the 54 matching Urbanized Areas/cities. The larger the angle between the visualized diagonal line and a virtual line between the axis intersection at 0/0 and a chosen point of interest (which represents one city), the larger is the relative discrepancy in density between OSM and the reference data for that city. The data plots indicate a large variation in relative OSM data completeness for trails and lanes. Visual comparison of Figure 5a and Figure 5b further suggests that points in the lanes plot have on average a larger deviation angle from the diagonal line than trail points. To quantify this we computed a trail ratio (OSM trail density / reference trail density) and lane ratio (OSM lane density / reference lane density) for each of the 54 cities. Cities with 0 km in the trail or lane denominator were excluded, which left 48 cities for comparison. A paired samples t-test on log-transformed trail and lane ratios revealed a significantly larger trail than lane ratio ($t=4.764, df=47, p<0.0001$, two-tailed). This means that OSM trails are generally more completely mapped than lanes when compared to the used reference dataset. A possible explanation is that trails have their own geometry independent of roads, whereas a bicycle lane is coded as a road attribute without its own geometry. Thus newly mapped trail features are visually more distinct than mapped lanes, which may result in a higher motivation for an OSM mapper to add bicycle trails rather than lanes.

A few cities have a higher trail or lane density in OSM than in the reference data set, as can be seen by the points located above the diagonal line in Figure 5. These values are underlined in the density columns in Table 1. Visual inspection of OSM trail and lane features in
the affected cities, and comparison with a background satellite image and the Google Maps Cycling layer reveal a few common reasons for such a situation:

a) Sidewalks or footpaths between buildings are in OSM mapped as trails (Figure 6a, Jacksonville, FL).

b) Paths with grass or dirt surface are mapped as trails (Figure 6b, Denver, CO), although they are not mapped in Google Maps and data repositories of local agencies.

c) Bicycle lanes are mapped on roads where there are none (Figure 6c, Memphis, CO)

d) Recently constructed bicycle lanes and trails have already been mapped in OSM, but they are not part of the 2008 reference dataset.

In general only few errors of commission (cases a-c) were identified in OSM data. Thus the higher density of OSM trail and lane data compared to the 2008 reference data is primarily caused by new trails and lanes that were constructed between 2008 and 2013 (case d).

4.2 Comparison of OSM and Google Bicycling Data Completeness
This section describes a manual validation approach that assesses the completeness of OSM and Google cycling features in 25 km² test areas located in the Central Business Districts of Portland and Miami (see Figure 3 and Figure 4). Completeness can be measured by 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 a bicycle feature exists in the real world but is not mapped. This error can occur in two ways: First, the geometry of the feature has been mapped but incorrectly tagged, i.e. not as a trail or lane. Second, the geometry is missing, i.e., the feature has not been mapped with any tag.
An error of commission occurs if a cycling feature is present in the dataset but not in the real world. This error can be caused by an incorrect geometry, meaning that a cycling feature is mapped although no linear feature of any type can be found in the real world at that specific location. The error can also be caused by incorrect tagging, which occurs if coding a different feature, such as a pedestrian-only footpath, as cycling feature, e.g. a bicycle trail.

Figure 7 gives examples of these errors in the London road network. Figure 7a shows in Google Street View a footpath that can also be traversed by bicycle, as indicated by the bicycle path sign on the pillar to the right. Figure 7b maps the situation as found in OSM, where the highlighted path corresponds to the segment photographed in the Google Street View image. Its tags correctly state that bicycles are allowed. As opposed to this, no geometry is shown for this path in Google Maps (Figure 7c), which is an error of omission caused by a missing geometry in the Google data set.

The second situation shows a street without a bicycle lane (Figure 7d). While Google Maps does correctly not map a bicycle lane on this road (Figure 7f), the OSM tags of this road indicate a bicycle lane (Figure 7e). The latter is therefore an error of commission in OSM caused by incorrect tagging of the road.

For the assessment of OSM and Google data completeness a reference data set with ground truth is needed. We compiled such a reference data set from three types of sources which are: (1) recent trail and lane data layers obtained from local public planning agencies in Portland and Miami, respectively, as described above; (2) OSM; and (3) the Google Maps Cycling layer. These three resources were merged into one layer and clipped to the 25 km² test areas. Further, each mapped feature was manually compared with a background image in ESRI’s ArcMap 10
and Google Maps for a potential error of commission, which was corrected if occurring. This approach provided a trail and lane reference dataset for both test areas.

Next the first author of this paper identified errors of omission by comparing OSM and Google trail and lane layers with the combined reference dataset. Further, errors of commission were identified for OSM and Google data through comparison of mapped features with aerial background imagery and Google Street View. The following four layers were checked: (1) OSM bicycle trails, (2) Google bicycle trails, (3) OSM designated lanes, (4) Google designated lanes. Table 2 summarizes the analysis results for the Portland and Miami test areas. For each city, the upper half refers to bicycle trails and the lower half to on-street lanes. The first “Total” value in each trail or lane section lists the total length of reference trails and lanes, respectively. This reference length consists of correctly mapped and omitted features for each data source.

The Portland test area contains about 26 km of trail features, out of which 86.4% were correctly mapped in OSM and 78.4% in Google Maps. In both data sources, a higher percentage of trail omission errors is caused by incorrect tags rather than by missing geometries. For designated lanes the reference length totals 59 km where in both data sources lanes are more completely mapped than trails. In both data sources errors of commission in trail and lane data were primarily caused by incorrect tags. Error of commission rates were at or below 7% for trail and lane layers except for OSM trails (error rate 22.0%) where footpaths were sometimes mapped as bicycle trails.

For the Miami test area the mapping of the 22 km trail features in OSM and Google Maps is less complete than for the Portland test area and reaches only 22.8% for OSM and 36.5% for Google. Most errors of omission in trails come from missing geometries, e.g., bicycle trails along rivers and in parks that were not mapped at all, i.e. not even as a pedestrian path. In this case the
reference data set from the Miami-Dade MPO provided a more complete map of cycling trails than both OSM and Google Maps did. Bicycle lanes in OSM are almost perfectly mapped while Google shows only 66.6% completeness. Errors of commission are lower than for Portland ranging between 0.1% and 7.7% for trails and lanes in both data sources. The observed difference in mapping completeness of OSM trail features between Portland (86.4%) and Miami (22.8%) is also reflected in Table 1 where Portland ranks 5th in trail density with only a small difference between the OSM trail density (0.476) and the 2008 reference density (0.558). As opposed to this, Miami only ranks 62nd in this category, revealing a much larger discrepancy between the OSM trail density (0.042) and the 2008 reference density (1.133). Thus the manual validation task supports findings related to Table 1 and Figure 5 which indicate that the OSM mapping completeness of bicycle trails and lanes varies between different regions.

5. DISCUSSION

Given the limited selection of road network datasets available that provide bicycle trail and lane information, OSM proves at least for a subset of analyzed Urbanized Areas in the US to be a relatively complete and easily accessible data source for basic analysis tasks. This research revealed differences in OSM data quality for different US cities (Table 1). A possible solution to resolve the low completeness of OSM features in affected areas could be to run a one-time data import from publicly available alternative data sources, such as MPO data. While data imports are discouraged within the OSM community since they could affect other contributors’ manual data collection and editing efforts (OSM, 2013a), one option to improve the quality of OSM cycling layers is to trace current layer files (e.g. shapefiles) that are provided by public agencies in OSM editors (Zielstra et al., 2013). With regards to our approach chosen for the assessment of OSM data completeness at the aggregated city level it must be mentioned that the comparison of
densities of trail and lane features between OSM and the reference dataset may not adequately evaluate spatial completeness, but that an analysis of overlapping features would provide a better insight instead. This was pointed out by Jackson et al. (2013) in the context of Points-Of-Interest completeness analysis. In their study, the authors compared the locations of mapped schools between three data sources in the larger Denver area, and while in the total numbers of schools were reasonable close in all three datasets (between 402 and 412), a closer assessment revealed that only 281 schools were common in all three datasets. Since we did not have access to the 2008 reference lane and trail data, but only to trail and lane densities for complete cities, we were limited to the comparison of densities and subsequent checks for errors of commission for the OSM dataset (end of section 4.1). However, to get a better understanding of where differences in densities between two compared data sets come from, we conducted a feature based comparison between OSM and Google Maps for parts of Miami and Portland, which revealed errors of commission and omission in both compared datasets (section 4.2).

This feature based comparison of the completeness between OSM and Google Maps trail and lane data did not identify a single best dataset. While OSM performed better than Google Maps in trail completeness for Portland and lane completeness for Miami, Google Maps provided more complete trail information for Miami. The varying completeness values of different data sources between different cities are consistent with what Cipeluch et al. (2010) found in a study where the spatial coverage of OSM road networks was visually compared to that of Google Maps and Bing Maps for five cities in Ireland. The authors manually counted road segments of different road types (e.g., motorways, national primary roads, roundabouts) for 4-square kilometer grids using a visualization in OpenLayers. OSM compared favorably in coverage for three cities, especially in those road categories where new road construction has
been recently completed. However, in one city OSM contained almost 50% fewer estate roads than either Google or Bing. Further it was found that OSM data collection proved to be problematic in housing access roads where security gates were present when compared to Google or Bing maps.

It should be noted that the analysis presented in this paper reflects the current state of collected OSM data. The OSM data quality is likely to improve further in the near future, not only in the US but also in other regions of the world, given the continuous data collection efforts. As an example, Figure 8 visualizes the development of total length (left column) and density (right column) of bicycle trail and lane features in OSM for seven US and European metropolitan areas between 2009 and 2013. Whereas European cities demonstrate particularly high growth rates in total length of mapped bicycle trails, the data growth is more balanced between US and European cities for designated lanes, with Portland standing out as a city with especially high data collection efforts. All seven cities, except for London, nearly doubled the total length of mapped trails between 2009 and 2013 or exceeded that rate. All seven cities also nearly doubled the total length of mapped bicycle lanes, or exceeded that rate.

6. CONCLUSIONS
This paper conducted a review of OSM trail and lane data for 78 Urbanized Areas in the US and their comparison with trail and lane densities from 2008 reference data. To the best of our knowledge this is the first study that assesses the completeness of OSM bicycle feature data for the US. It is a current topic given the rapidly growing bicycle infrastructure in the US and the fact that the proprietary data providers do not include information about bicycle lanes and trails in their vector datasets that are available for purchase. The analysis demonstrated a large variation in OSM data completeness between cities. Some cities showed even higher trail and
lane supply in OSM than the 2008 reference data which is primarily because of active mappers collecting and mapping newly constructed bicycle facilities. In areas that reveal low OSM data completeness a comparison with Google Maps Bicycling or governmental trail and lane feature layers can be helpful to detect missing trail and lane supply. However, licensing terms and conditions may prevent conflation efforts towards a more sophisticated dataset for cyclists, e.g., when considering a combination of OSM and Google Maps datasets. On the one side when creating derived work from Google Maps data, e.g. through digitizing roads, such data retains copyright conditions of the original and is subject to license fees, which OpenStreetMap tries to avoid by encouraging independent data collection. On the other side the current licensing conditions of the OSM dataset specifically note that any potential data conflation is only allowed if the newly created dataset is available under the same open (ODbL) license, resulting in a different set of limitations for interested users (e.g. commercial data providers). As opposed to this data conflation between freely available governmental data (e.g. TIGER/Line) and OSM is possible (Zielstra et al., 2013).

While trail and lane features, which were the focus of this study, are important for various transportation planning and analysis tasks, additional road attributes, such as traffic volume, and facility features, such as undesignated lanes, are necessary to compute other bicycle related analysis measures, such as the Bicycle Level of Service or Bicycle Compatibility Index. However, some of these attributes and feature types are currently not supported in the OSM coding scheme (Hillsman and Barbeau, 2011). One aspect of future work is therefore to observe the likely development of additional bicycle related attributes and feature categories in the OSM coding schemes. Another aspect of future work involves the analysis of OSM data completeness
for rural areas, which has so far been only conducted for streets with motorized traffic (Zielstra and Hochmair, 2011b; Zielstra and Zipf, 2010), but not for cycling infrastructure.
References
Barnes G and Krizek K 2005 Estimating Bicycling Demand. Transportation Research Record: Journal of the Transportation Research Board 1939: 45-51
Bennett J 2010 OpenStreetMap. Birmingham, Packt Publishing
Hillsman E L and Barbeau S J 2011 Enabling Cost-Effective Multimodal Trip Planners through Open Transit Data. Tallahassee, FL. Prepared for Florida Department of Transportation.


Mooney P and Corcoran P 2012b How social is OpenStreetMap? In Proceedings of the 15th AGILE International Conference on Geographic Information Science. Avignon, France


OSM 2013c OSM wiki for Map Features. WWW document, [http://wiki.openstreetmap.org/wiki/Map_Features](http://wiki.openstreetmap.org/wiki/Map_Features)


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


Zielstra D and Hochmair H H 2012 Using Free and Proprietary Data to Compare Shortest-Path Lengths for Effective Pedestrian Routing in Street Networks. *Transportation Research Record: Journal of the Transportation Research Board* 2299: 41-47

Zielstra D and Zipf A 2010 OpenStreetMap Data Quality Research in Germany. In *Proceedings of the Sixth international conference on Geographic Information Science (GIScience)*. Zurich, Switzerland
### Table 1 OSM Bicycle trails and designated lanes for selected Urbanized Areas in the United States

<table>
<thead>
<tr>
<th>Urbanized Area</th>
<th>POP '10 (mio.)</th>
<th>Total km</th>
<th>Trail Density</th>
<th>Ref. Dens.</th>
<th>Rank</th>
<th>Total km</th>
<th>Lane Density</th>
<th>Ref. Dens.</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omaha, NE-IA</td>
<td>0.725</td>
<td>658</td>
<td>0.938</td>
<td>1.435</td>
<td>1</td>
<td>5</td>
<td>0.007</td>
<td>0.009</td>
<td>50</td>
</tr>
<tr>
<td>Denver-Aurora, CO</td>
<td>2.374</td>
<td>1456</td>
<td>0.342</td>
<td>0.404</td>
<td>2</td>
<td>87</td>
<td>0.050</td>
<td>0.170</td>
<td>27</td>
</tr>
<tr>
<td>Minneapolis-St. Paul, MN-WI</td>
<td>2.651</td>
<td>1452</td>
<td>0.549</td>
<td>1.523</td>
<td>3</td>
<td>176</td>
<td>0.067</td>
<td>0.738</td>
<td>22</td>
</tr>
<tr>
<td>Colorado Springs, CO</td>
<td>0.559</td>
<td>265</td>
<td>0.545</td>
<td>0.635</td>
<td>4</td>
<td>0</td>
<td>0.000</td>
<td>0.415</td>
<td>70</td>
</tr>
<tr>
<td>Portland, OR-WA</td>
<td>1.850</td>
<td>646</td>
<td>0.476</td>
<td>0.558</td>
<td>5</td>
<td>1425</td>
<td>1.049</td>
<td>1.310</td>
<td>1</td>
</tr>
<tr>
<td>Washington, DC-VA-MD</td>
<td>4.587</td>
<td>1293</td>
<td>0.378</td>
<td>0.912</td>
<td>6</td>
<td>188</td>
<td>0.055</td>
<td>0.728</td>
<td>25</td>
</tr>
<tr>
<td>Chicago, IL-IN</td>
<td>8.608</td>
<td>2192</td>
<td>0.347</td>
<td>0.181</td>
<td>7</td>
<td>332</td>
<td>0.053</td>
<td>0.502</td>
<td>26</td>
</tr>
<tr>
<td>Albuquerque, NM</td>
<td>0.741</td>
<td>221</td>
<td>0.340</td>
<td>0.747</td>
<td>8</td>
<td>284</td>
<td>0.438</td>
<td>0.913</td>
<td>4</td>
</tr>
<tr>
<td>Concord, CA</td>
<td>0.616</td>
<td>174</td>
<td>0.330</td>
<td></td>
<td>9</td>
<td>150</td>
<td>0.284</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Mission Viejo-L. Forest-S. Clem., CA</td>
<td>0.584</td>
<td>123</td>
<td>0.316</td>
<td></td>
<td>10</td>
<td>181</td>
<td>0.464</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Memphis, TN-MS-AR</td>
<td>1.060</td>
<td>19</td>
<td>0.015</td>
<td>0.050</td>
<td>69</td>
<td>34</td>
<td>0.027</td>
<td>0.007</td>
<td>31</td>
</tr>
<tr>
<td>Baton Rouge, LA</td>
<td>0.594</td>
<td>11</td>
<td>0.012</td>
<td>0.260</td>
<td>70</td>
<td>3</td>
<td>0.003</td>
<td>0.213</td>
<td>60</td>
</tr>
<tr>
<td>Charlotte, NC-SC</td>
<td>1.249</td>
<td>17</td>
<td>0.009</td>
<td>0.124</td>
<td>71</td>
<td>4</td>
<td>0.002</td>
<td>0.231</td>
<td>64</td>
</tr>
<tr>
<td>Nashville-Davidson, TN</td>
<td>0.970</td>
<td>13</td>
<td>0.009</td>
<td>0.099</td>
<td>72</td>
<td>2</td>
<td>0.001</td>
<td>0.065</td>
<td>69</td>
</tr>
<tr>
<td>Richmond, VA</td>
<td>0.954</td>
<td>11</td>
<td>0.009</td>
<td></td>
<td>73</td>
<td>0</td>
<td>0.000</td>
<td></td>
<td>75</td>
</tr>
<tr>
<td>Bridgeport-Stamford, CT-NY</td>
<td>0.923</td>
<td>9</td>
<td>0.007</td>
<td></td>
<td>74</td>
<td>0</td>
<td>0.000</td>
<td></td>
<td>76</td>
</tr>
<tr>
<td>Knoxville, TN</td>
<td>0.559</td>
<td>7</td>
<td>0.006</td>
<td></td>
<td>75</td>
<td>0</td>
<td>0.000</td>
<td></td>
<td>77</td>
</tr>
<tr>
<td>Birmingham, AL</td>
<td>0.749</td>
<td>6</td>
<td>0.005</td>
<td>0.010</td>
<td>76</td>
<td>2</td>
<td>0.001</td>
<td>0.013</td>
<td>66</td>
</tr>
<tr>
<td>Urban Honolulu, HI</td>
<td>0.802</td>
<td>0</td>
<td>0.000</td>
<td>0.548</td>
<td>77</td>
<td>26</td>
<td>0.059</td>
<td>0.513</td>
<td>24</td>
</tr>
<tr>
<td>McAllen, TX</td>
<td>0.729</td>
<td>0</td>
<td>0.000</td>
<td></td>
<td>78</td>
<td>0</td>
<td>0.000</td>
<td></td>
<td>78</td>
</tr>
</tbody>
</table>
### Portland

<table>
<thead>
<tr>
<th>TRAILS</th>
<th>OSM</th>
<th>Google</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Error of omission</td>
<td>Error of omission</td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>Tag</td>
</tr>
<tr>
<td></td>
<td>[m]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22605</td>
<td>3008</td>
</tr>
<tr>
<td></td>
<td>86.4</td>
<td>11.5</td>
</tr>
<tr>
<td></td>
<td>Error of commission</td>
<td>Error of commission</td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>Tag</td>
</tr>
<tr>
<td></td>
<td>[m]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5216</td>
<td>552</td>
</tr>
<tr>
<td></td>
<td>19.9</td>
<td>2.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LANES</th>
<th>OSM</th>
<th>Google</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Error of omission</td>
<td>Error of omission</td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>Tag</td>
</tr>
<tr>
<td></td>
<td>[m]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>56146</td>
<td>2796</td>
</tr>
<tr>
<td></td>
<td>95.3</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>Error of commission</td>
<td>Error of commission</td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>Tag</td>
</tr>
<tr>
<td></td>
<td>[m]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1672</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>0.0</td>
</tr>
</tbody>
</table>

### Miami

<table>
<thead>
<tr>
<th>TRAILS</th>
<th>OSM</th>
<th>Google</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Error of omission</td>
<td>Error of omission</td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>Tag</td>
</tr>
<tr>
<td></td>
<td>[m]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4960</td>
<td>5162</td>
</tr>
<tr>
<td></td>
<td>22.8</td>
<td>23.8</td>
</tr>
<tr>
<td></td>
<td>Error of commission</td>
<td>Error of commission</td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>Tag</td>
</tr>
<tr>
<td></td>
<td>[m]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>311</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.4</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LANES</th>
<th>OSM</th>
<th>Google</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Error of omission</td>
<td>Error of omission</td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>Tag</td>
</tr>
<tr>
<td></td>
<td>[m]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9853</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>99.2</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Error of commission</td>
<td>Error of commission</td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>Tag</td>
</tr>
<tr>
<td></td>
<td>[m]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0</td>
</tr>
</tbody>
</table>

25
FIGURES

Bicycle trails:

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');

Figure 1 SQL queries to extract bicycle trails and designated lanes
Figure 2 Bicycle trails and designated lanes; street images are taken from Google Street View; symbolic illustrations to the left are taken from the OSM wiki for bicycle information (OSM, 2013b)
Figure 3 OSM bicycle trails and designated lanes in the Portland Urbanized Area
Figure 4 OSM bicycle trails and designated lanes in the Miami Urbanized Area
a) Figure 5 OSM trail vs. reference trail densities (a) and OSM lane vs. reference lane densities (b)
Figure 6 Explaining high OSM trail and lane densities: Mapping sidewalk as bicycle trail (a); mapping grass path as bicycle trail (b); mapping bicycle lane when there is none (c)
Error of omission:

highway=footway + bicycle=yes

a) Google Street View  b) OSM (correct)  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 (correct)

Figure 7 Error of omission and error of commission
Figure 8 Growth of OSM cycling data between 2009 and 2013