Transactions in GIS

Research Article Transactions in GIS, 2016, 20(6): 925–947

User Contribution Patterns and Completeness Evaluation of Mapillary, a Crowdsourced Street Level Photo Service

Levente Juhász and Hartwig H. Hochmair

Geomatics Program, University of Florida

Abstract

Mapillary is a Web 2.0 application which allows users to contribute crowdsourced street level photographs from all over the world. In the first part of the analysis this article reviews Mapillary data growth for continents and countries as well as the contribution behavior of individual mappers, such as the number of days of active mapping. In the second part of the analysis the study assesses Mapillary data completeness relative to a reference road network dataset at the country level. In addition, a more detailed completeness analysis is conducted for selected urban and rural areas in the US and part of northern Europe for which the completeness of Mapillary data will also be compared with that of Google Street View. Results show that Street View provides generally a better coverage on almost all road categories with some exceptions for pedestrian and cycle paths in selected cities. However, Mapillary data can be conveniently collected from any mobile device that is equipped with a photo camera. This gives Mapillary the potential to reach better coverage along off-road segments than Google Street View.

1 Introduction

Geolocated street-level photographs are an important data source for a variety of transportation analysis tasks, including the identification of road features, such as traffic signs (Gonzalez et al. 2014), the evaluation of wheelchair accessibility of sidewalks (Hara et al. 2014), or the deployment of navigational tools for visually impaired citizens (Guy and Truong 2012). Mapillary is the first platform to provide detailed street photos based on crowdsourcing, adding to the list of Web 2.0 applications that administer and facilitate the collection of Volunteered Geographic Information (VGI) (Goodchild 2007). Mapillary started its public service in early 2014 and is run by a company located in Malmö, Sweden, with a satellite office in Los Angeles, California. Imagery is provided at http://www.mapillary.com under the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) License. This means that everyone is free to use the data, even for commercial purposes.

Mapillary, like other VGI data sources, can be seen in different contexts (Haklay 2013). Its foremost aim is to produce geographic information, which necessitates consideration of spatial data quality (Goodchild 2007). With data completeness being one of the major data quality elements, data collection and data quality are closely intertwined in Mapillary. In this context, this study analyzes data growth and completeness on a worldwide scale as well as for selected local areas. In the context of user participation (Elwood et al. 2012) it will assess user loyalty

Address for correspondence: Levente Juhász, University of Florida, Geomatics Program, 3205 College Avenue, Ft. Lauderdale, FL-33314, USA. E-mail: levente.juhasz@ufl.edu

 doi: 10.1111/tgis.12190

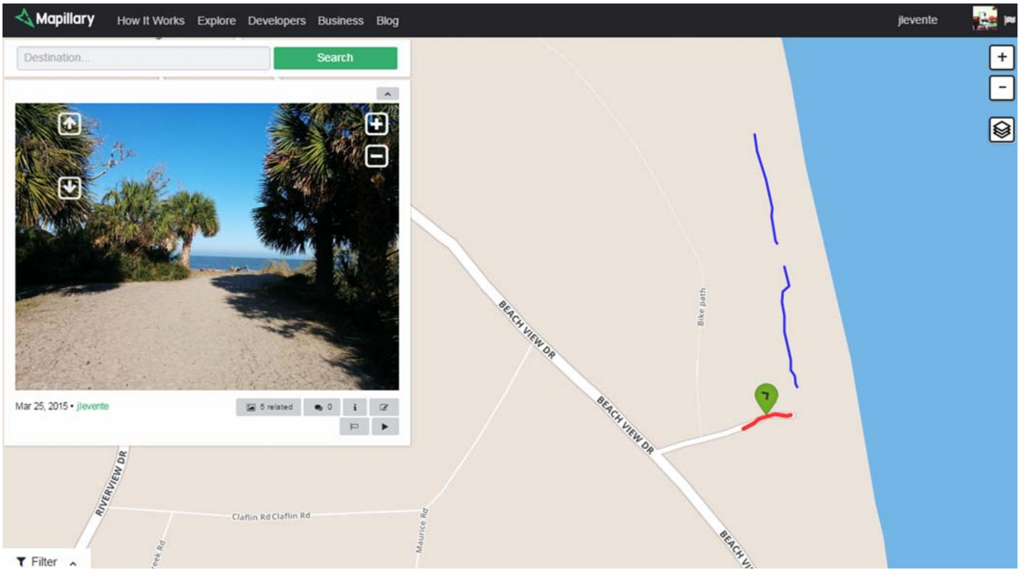

Figure 1 Mapillary image and mapped GPS traces as seen on the website

in data contribution. We assume that the Mapillary's ShareAlike license attracts mostly contributors who support the idea of open data over proprietary data, even if proprietary data, or part of them, are provided free-of-charge, as is the case with Google Street View data (from here on simply called Street View). The fact that Mapillary imagery has already been collected in all five continents of the world renders Mapillary data a useful supplemental data source to commercial products of street-level photographs, such as Street View. For selected urban and rural areas, therefore, this study will also compare data completeness between Mapillary and Street View relative to reference road datasets along different road categories. Street View provides street level photographs as 360° panorama images on selected roads from all over the world. The images are primarily taken from cars equipped with professional cameras. However, for selected bicycle and hiking paths Street View images are captured through specifically developed camera backpacks and tricycles with camera mounts.

Mapillary data contributors can upload photos using a smartphone application, or manually on the Mapillary website. Video upload functionality was added in April 2015, which automatically extracts geotagged photos from the video stream. Videos need to be accompanied by a GPS log for geocoding purposes. Since Mapillary facilitates data capture from any GPS enabled mobile device equipped with a camera it increases the range of potential data contributors compared to professional camera teams employed by Google for Street View data collection. This flexibility results in Mapillary imagery also being frequency collected from off-road paths, such as hiking trails.

Figure 1 shows how a captured street level photograph can be viewed on the Mapillary website. The blue lines on the map indicate all GPS sequences within the area along which photos were taken, and the green marker with the arrow along the red GPS sequence shows the image location together with the orientation of the camera. For devices without a built-in compass, the camera orientation is determined by the travel trajectory, using the current location and the location of the next photo of the sequence. The orientation can be manually edited later.

The remainder of the article is structured as follows. The next section reviews related research, which is followed by a description of the data sources and the applied analysis methods in Section 3. Section 4 presents results of the contribution analysis, followed by conclusions and directions for future work in Section 5.

2 Related Work

The availability of street level photographs can eliminate time and resource consuming in-person fieldwork. Identification of urban neighborhood features is commonly used in health research since the characteristics of the built environment influence health behavior. Clarke et al. (2010) gathered neighborhood measures (e.g. park, playground, graffiti presence) from Street View imagery to characterize communities. Rundle et al. (2011) and Griew et al. (2013) found that Street View can be used to audit neighborhood environments by matching results obtained from Street View imagery data to data collected on-site. However, on-site assessments should be conducted when characterizing micro-environments for special activities, such as cycling (Vanwolleghem et al. 2014). Several studies applied computer vision to Street View imagery, for example, to identify sidewalk curbs (Hara et al. 2014), or to determine the geographic location of photographs based on reference imagery extracted from Street View (Zamir and Shah 2010). Yin et al. (2015) provide technical details on downloading and assembling Street View images and their use for pedestrian detection using machine vision and learning technology. Guy and Truong (2012) developed a navigation prototype that relies on the manual identification of road intersection features from Street View images to augment visually impaired pedestrians' sensory information with a richer depiction of the environment. Despite the wide spread use of Street View imagery in geospatial applications and research, the spatial coverage of Street View, i.e. data completeness and geographic extent, has so far not been discussed in the literature. An earlier version of the study presented in this article revealed that Street View provides fairly complete coverage where service is offered in a city (Juhász and Hochmair 2015). For the 20 largest German cities Street View covers 77–92% of the main roads, 58–84% of the residential roads, but fewer than 4% of off-road pedestrian and bicycle network segments. No statistical relationship could be identified between the presence of Street View data and the Mapillary completeness value for the 40 German cities analyzed. This implies that data collection of Mapillary is driven by factors other than the prior existence of proprietary data, such as contributors' idealistic attitude to contribute collected data even in areas where proprietary data are provided free of charge. A related study compared the data quality of mapped roads between Google Maps (not Street View, though), Bing Maps and OpenStreetMap (OSM) in Ireland where spatial coverage, currency and ground-truth positional accuracy were measured. The authors found that the performance of the three datasets varied between the five analyzed cities and that there was no clear best dataset (Cipeluch et al. 2010).

Part of the completeness analysis conducted in this article relies on reference data, where OSM is used as one of the data sources. Although the data quality of OSM is not assured by standards or an authoritative agency, several studies report that OSM provides high road coverage, especially in urban areas, rendering it a viable free alternative to commercial datasets (Girres and Touya 2010; Haklay 2010; Zielstra and Zipf 2010). Other research studies assessed OSM data completeness on off-road paths, e.g. pedestrian segments (Zielstra and Hochmair 2011) and bicycle trails (Hochmair et al. 2015). Results of these studies indicate that OSM provides a robust reference dataset for determining Mapillary and Street View data completeness even for off-road data analysis.

Several aspects of data analysis applied on the Mapillary dataset presented in this article have been previously applied on other VGI data sets as well as non-spatial user generated content. For example, participation inequality is well-known in online social networks and in large-scale, multi-user communities (Nielsen 2006). It means that most of the data is contributed only by a small fraction of users. Participation inequality could be observed for Wikipedia (Javanmardi et al. 2009), OSM (Neis and Zipf 2012), dronestagram (Hochmair and Zielstra 2015), and Flickr and Twitter (Li et al. 2013). Techniques to identify a contributor's home region from data contribution patterns or social interactions with other users have been developed for OSM (Neis and Zipf 2012; Zielstra et al. 2014), Twitter (Davis Jr et al. 2011), and Facebook (Backstrom et al. 2010). The extent of human mobility patterns, measured as the radius of gyration, have been derived from Foursquare check-ins (Cheng et al. 2011), tweets (Hawelka et al. 2014) and mobile phone records (González et al. 2008). The localness of user generated content, i.e. the share of content that is contributed from within 100 km of a contributor's specified home location, was found to be higher in Flickr than in Wikipedia (Hecht and Gergle 2010).

3 Study Set-up

3.1 Overview of the Analysis

The study analyzes Mapillary data that were collected within the first year since the inception of the Mapillary project. The first part of the analysis focuses on measuring the aggregated data volume over time, for both countries and continents, as well as determining contribution patterns of individual users. The second part of the analysis examines Mapillary data completeness relative to reference road datasets from OpenStreetMap and Esri, both at the country and city level. It compares also Mapillary to Street View completeness for selected cities in the US and Europe.

3.2 Tile System

Whereas measuring Mapillary data contributions for the first part of the analysis relies on vector data operations, the second part of the analysis, which addresses data completeness, requires overlaying road data from several sources, i.e. Mapillary, Google, OSM, and Esri. For this purpose we used raster tiles of the Web Map Tile Service (WMTS) (Masó et al. 2010) as a common spatial reference system. This so called XYZ tile scheme, which was first used by Google (2015), became a de facto standard in Web mapping and is used by numerous other map providers as well, including Bing Maps, Yahoo Maps, OSM and Mapbox. Tiles are provided as 256 x 256 pixel images. For the visualization of geographic data, the world is divided into tiles corresponding to zoom levels. In each zoom level, tiles are indexed by X (column) and Y (row) values starting from the top-left corner (Figure 2). Zoom level 0 covers the whole world in one tile. The total number of tiles at a zoom level z is 2^{2z}. Logically, the system can be considered to be a hierarchy of folders and files, where each zoom level is a folder, each X coordinate is a subfolder and each Y coordinate is a raster image file. Geographic coordinates can be converted to tile coordinates to determine in which tile a location on the Earth's surface can be found (Equation 1a) and tile coordinates can also be converted to geographic coordinates indicating the top-left corner of a tile (Equation 1b) as follows:

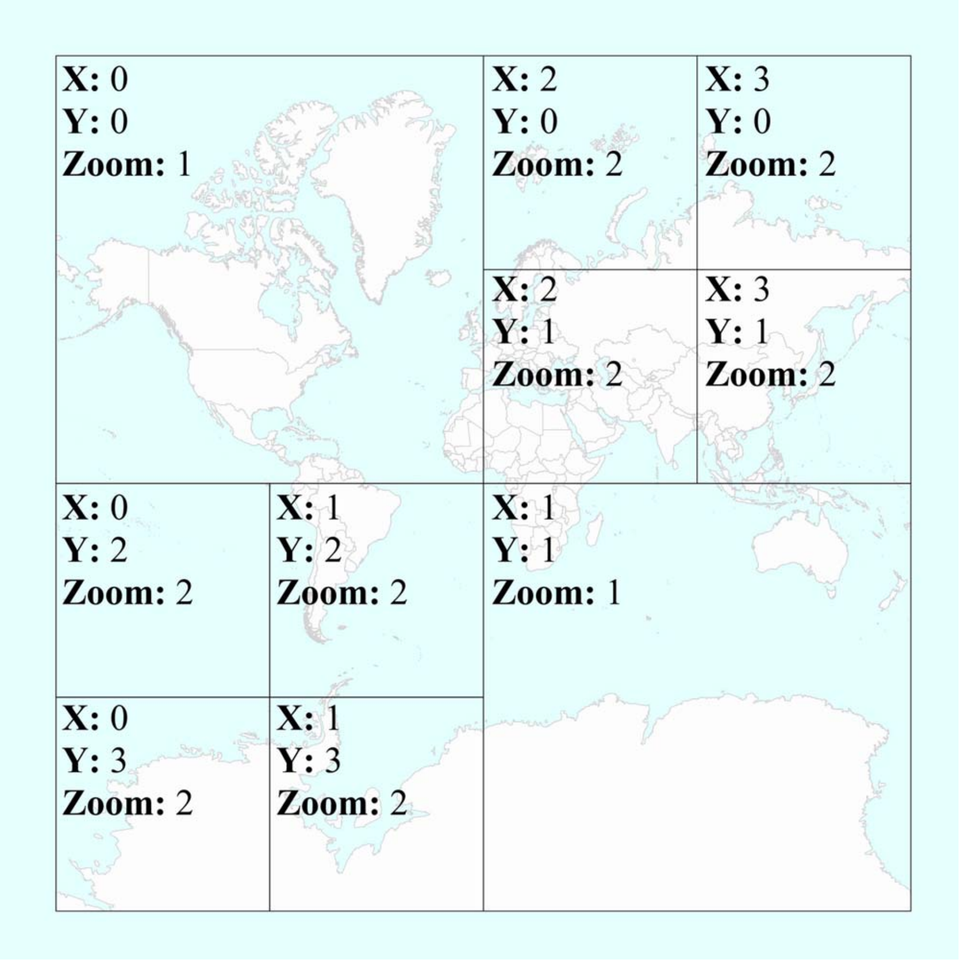

Figure 2 Zoom levels with tile coordinates

$$X = \left\lfloor \frac{lon+180}{360} \cdot 2^z \right\rfloor$$

$$Y = \left\lfloor \left(1 - \frac{1}{\pi} \cdot ln \left(\tan\left(lat \cdot \frac{\pi}{180} \right) + \frac{1}{\cos\left(lat \cdot \frac{\pi}{180} \right)} \right) \right) \right\rfloor \cdot 2^{2z} \tag{1a}$$

$$lon = \frac{X}{2^z} \cdot 360 - 180$$

$$lat = \arctan\left(\sinh\left(\pi - \frac{Y}{2^z} \cdot 2\pi \right) \right) \cdot \frac{180}{\pi} \tag{1b}$$

where X, Y are the tile coordinates; z is the zoom level; lon, lat are the geographic coordinates) (OSM 2015).

In this research the XYZ tile scheme at zoom level 13 is used as a reference grid system for comparison and completeness analyses. At this zoom level the ground pixel size corresponds to approximately 19 m and tiles cover an area of approximately 4.9 km $\times$ 4.9 km. This setting conceals GPS positioning errors occurring during data collection of Mapillary, and avoids double counting of sequences taken along the same road. As a preliminary step, all data used in the completeness calculation were converted into this schema to make them comparable.

3.3 Data Description and Extraction

3.3.1 Mapillary

Mapillary provides access to its data in several ways. The Website visualizes the position of individual photos (as points) and sequences of photos (as lines). It allows viewing and downloading individual images associated with these points and lines (compare Figure 1). It can also filter uploaded image content by date and season and is able to show traffic signs that were extracted through image processing.

Mapillary provides also a JSON API which allows a user to search for images and sequences. Data can be downloaded in JSON and GeoJSON formats. These objects contain information about the photos and sequences, as well as the URL of corresponding photos. Mapillary data can be embedded in other applications and web mapping frameworks. For example, Mapillary photos and sequences have been included in the official OpenStreetMap iD editor since October 2014, making it easier for voluntary mappers to use visual Mapillary information for mapping purposes.

For this study, image sequences (as opposed to individual photos) were chosen for analysis since they provide a suitable representation of network segments that were mapped with Mapillary data. A sequence is a list of GPS nodes where each node represents the location of an image. Although sequences can be downloaded from the API, the Mapillary developer team provided us with a database dump, from 3 February 2015, which contained additional data information, such as a unique user ID for each sequence, timestamps showing when sequences were taken and uploaded to the site, and the geometry as LineStrings. This data was stored in a PostgreSQL database with the PostGIS extension enabled.

To avoid blurry images, the Mapillary smartphone application prevents the device from taking a photo if the phone is shaking. The same occurs during a loss of the GPS signal. In both cases, a sequence is automatically closed, and a new one started when an image is taken again after the interruption. A sequence is also interrupted if the distance between the current and the previous photo exceeds a certain threshold. Earlier versions of the application did not close a sequence in any of these cases, which led to long straight segments in the Mapillary dataset. Hence, for data analysis, straight segments of 1 km length or more were first removed, as shown by dashed lines in Figure 3. This ensured that sequences represented the true coverage of the taken imagery. For the completeness analysis, raster tiles were generated at zoom level 13 with Mapnik.

3.3.2 Google Street View

To be able to compare Street View to Mapillary coverage and to calculate Street View completeness, raster tiles of zoom level 13 are needed. Street View coverage tiles are highly generalized by default at zoom level 13 (Figure 4a) and can therefore not be used for comparison with other road networks. To make Street View tiles comparable to Mapillary tiles, Street View tiles were regenerated (Figure 4b). To do so, a vector version of Street View lines was extracted as

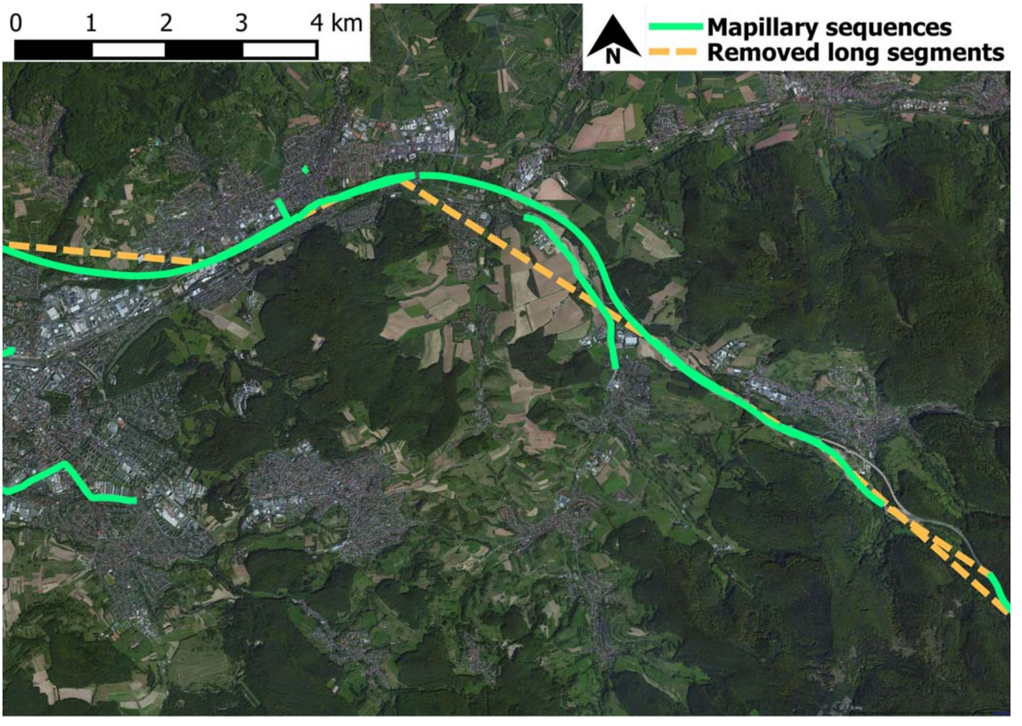

Figure 3 Removed long segments

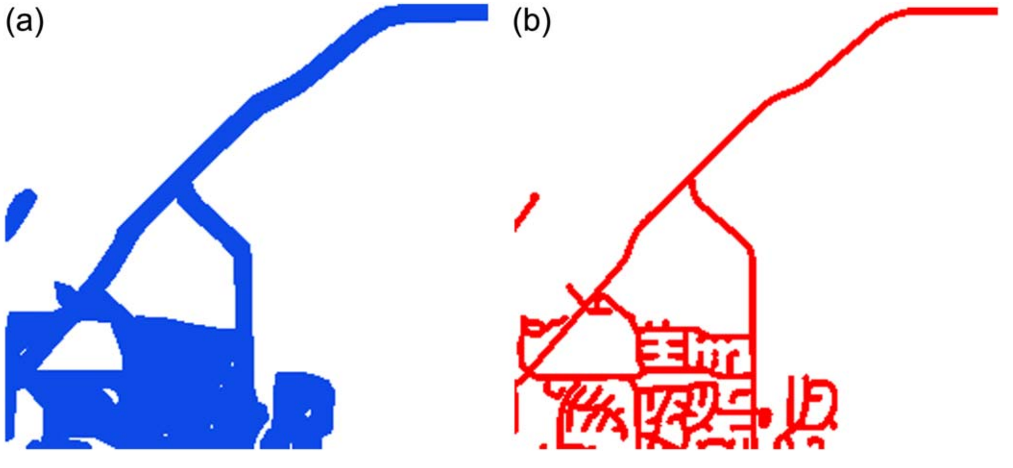

Figure 4 Street View tile (Zoom: 13): (a) Original; and (b) regenerated

follows. A self-developed client-side script downloaded Street View coverage tiles in PNG format at zoom level 17 for our areas of interest into the Web browser's cache. Individual images were then extracted from the cache and stored within the XYZ tile scheme. The ground pixel resolution at this level is 1.2 m, which provides a detailed representation of Street View coverage. After geocoding each tile using Equation 1b, PNG tiles were loaded into a GRASS GIS workspace where they were vectorized and uploaded into a PostgreSQL table with line geometries. The workflow of extracting line features used the algorithms of *r.patch*, *r.thin*, *r.to.vect* and *v.generalize* (GRASS Development Team 2015). Finally, zoom level 13 raster tiles were rendered with Mapnik to match Mapillary tiles.

3.3.3 Reference data for determining completeness

To calculate completeness of Mapillary and Street View data, reference datasets are needed. This study evaluates completeness at local and global scales. For the worldwide assessment of Mapillary data completeness, we used the Esri World Roads dataset, which is based on the DeLorme World Base Map. It can be assumed that this data product applies the same quality standard for the whole globe and that it therefore provides a more consistent reference dataset than OSM, which varies by region (Neis et al. 2013). The worldwide completeness analysis was conducted on main roads only, and therefore ferries and local roads were excluded from the reference dataset before the analysis.

A more detailed completeness analysis in selected areas of the US and Europe was conducted both for Mapillary and Street View. For this purpose an OSM database dump from February 2015 was downloaded from Geofabrik (http://download.geofabrik.de), which provides a comparable data quality for all selected areas. All roads were extracted with the Osmosis software tool using a *highway=** filter and uploaded into a PostgreSQL database. Additional queries were designed to extract the following road categories:

1. Main roads: connect settlements and cities;
2. Residential roads: minor, lower level roads with moderate traffic; and
3. Pedestrian/Cycle paths: minor elements of the road network used by pedestrians and cyclists for daily routine or recreational purposes.

Inaccessible roads, sidewalks, road crossings, tunnels and indoor features were excluded from the selection based on their tags. Visual inspection of the results showed a high number of pedestrian/cycle features close to higher road categories. We eliminated those pedestrian/cycle features that were within 25 m of the other two categories to analyze only off-road pedestrian/cycle features. Map tiles for zoom level 13 were generated from the Esri and OSM road datasets for the different geographic regions (global or local, respectively) and were subsequently used as reference datasets for the completeness analysis.

3.4 Completeness Computation for Mapillary and Street View

For the completeness analysis at the local scale we aimed to determine the percentage of reference roads, separated by road category, which were also mapped in Mapillary and Street View, respectively. To compute completeness, a self-developed python script compared the content of tiles of reference datasets with that of Mapillary and Street View tiles. More specifically, the script loaded each reference tile, identified pixels of reference roads and also counted the number of pixels overlapping with pixels in the Mapillary and Street View tiles. Results were loaded into a PostgreSQL database with polygon geometries.

The geographic extent of a tile at zoom level 13 is $\sim$4.9 $\times$ 4.9 km. For worldwide completeness, analysis of Mapillary data it was necessary to assign a mapped road segment to the correct country, which needed special data handling if a border between two countries was running across a tile. Affected tiles were subsequently divided into a refined grid system with a tile size of approximately 40 x 40 m. Pixel count values of the original tiles (green polygons in Figure 5) and the refined tiles, considering only those that are completely contained within a

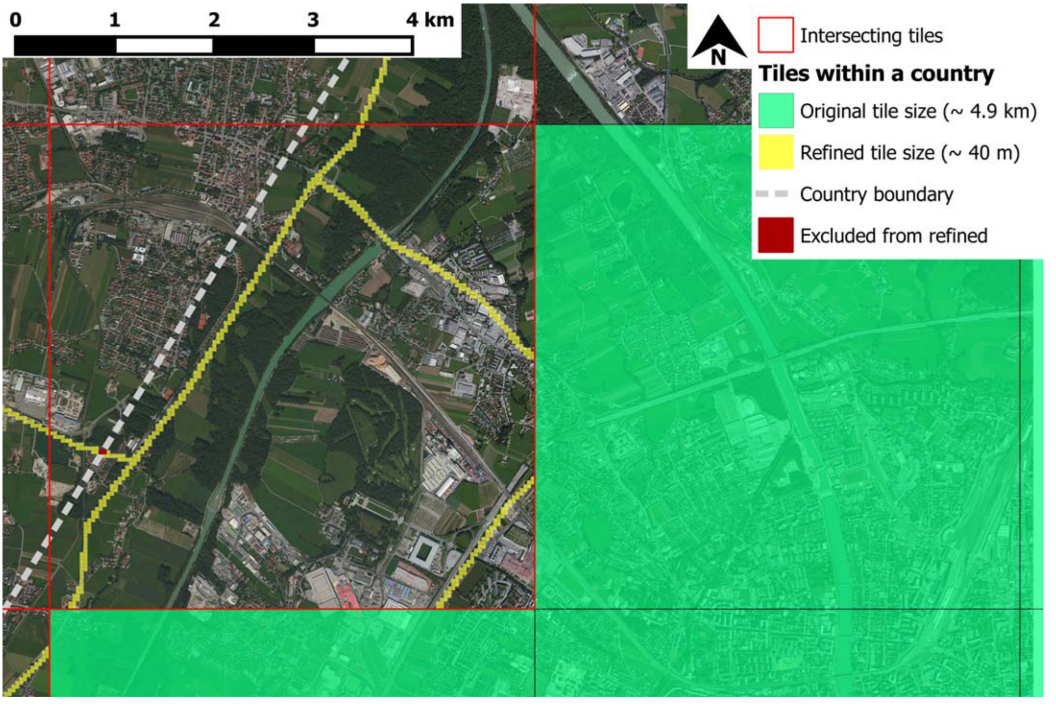

Figure 5 Refined tiles around country borders

country (yellow polygons in Figure 5), were assigned to the corresponding country. Completeness can then be calculated for administrative units using reference data as follows:

$$C = \frac{\sum Px}{\sum Ref} \tag{2a}$$

where C is the completeness of Mapillary or Street View, Px is a Mapillary or Street View pixel overlapping with a pixel from the reference road dataset, and Ref is a pixel from the reference road dataset. In addition, the relative completeness difference was calculated for selected tiles as part of the local analysis to compare Mapillary completeness with that of Street View on specific road categories as follows:

$$d(SV_r,\ Map_r) = \begin{cases} 0, & if\ \ SV_r + Map_r = 0 \\ \dfrac{(SV_r - Map_r)}{(SV_r + \ Map_r)}, & if\ \ SV_r + Map_r\ \ > 0 \end{cases} \tag{2b}$$

where d is the relative completeness difference, SV_r is the count of Street View pixels and Map_r is the count of Mapillary pixels on road category r. Relative completeness difference values range between 21 and 1. A positive value means that a tile contains more roads mapped with Street View than with Mapillary, and a negative value means the opposite. A value of zero can mean either that Street View and Mapillary have identical coverage, or that they have no coverage at all. Pixels were in the latter case excluded from the corresponding map in Figure 13.

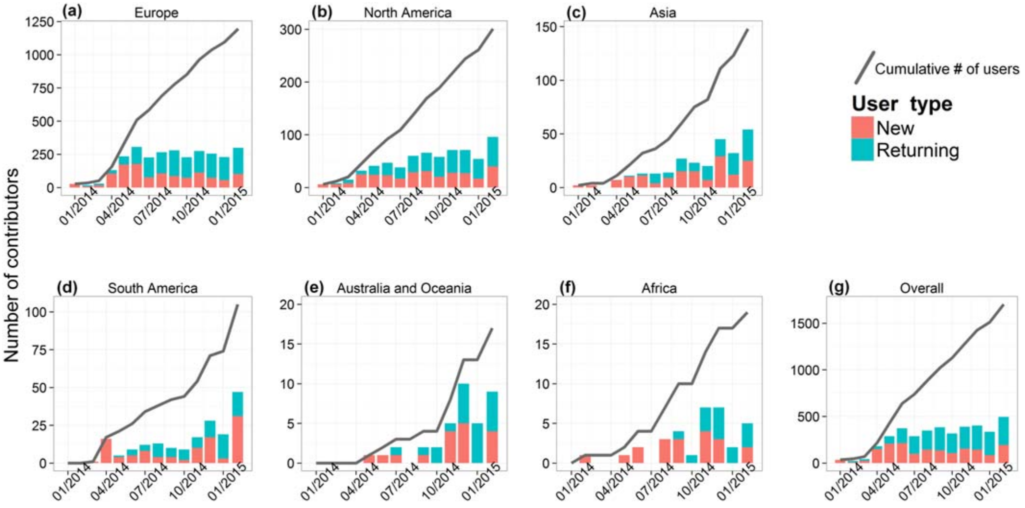

Figure 6 Number of users per continent contributing to Mapillary

4 Results

Section 4.1 reviews contribution patters on a larger scale as well as on the individual level. Section 4.2 analyzes global data completeness on main roads and compares the extent of Mapillary imagery to Street View at the city level.

4.1 Contribution Patterns of Mapillary Data

4.1.1 Country and continental level

Figure 6 shows for each month the number of users actively contributing to Mapillary, separated by continent (Figures 6a-f) and worldwide (Figure 6g). Each bar chart indicates the proportion of new and returning (those who mapped before) users. This split reveals that the majority of mappers contributes on a regular basis. Up to 3 February, 2015, data were contributed from 1,709 users worldwide (Figure 6g). The number of users is largest in Europe (1,194 users, Figure 6a), followed by North America (303 users, Figure 6b). In these two continents Mapillary users are more committed to Mapillary than in other continents, as can be seen by the higher portion of returning users each month.

Mappers covered GPS tracks of more than 209,000 km worldwide, where it should be noted that this number counts also overlapping sequences that were taken on the same road. The growth pattern for the total mapped distance is similar to that of user numbers. Again, Europe, followed by North America, are the most mapped continents (Figure 7a). Figure 7b shows the evolution of contributed data for the five most mapped countries. Mapillary data collection focused primarily on Germany, Sweden and Poland in Europe and on the US and Canada in North America. In addition, Figure 7 shows the evolution of the average distance mapped per user for continents (Figure 7c) and the same countries as before (Figure 7d). Interestingly, the average contribution per user is higher in 17 other countries than in the most mapped country, Germany. The countries with the highest average distance mapped per user value are Azerbaijan (561 km, one user), followed by Nicaragua (524 km, six users) and

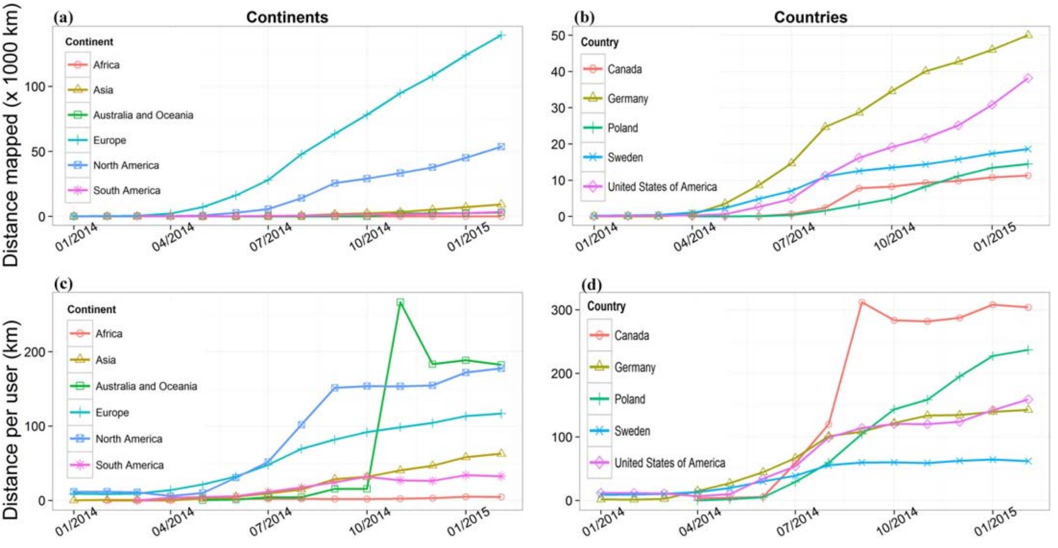

Figure 7 Total distance mapped per continent (a) and country (b), and average distance mapped per user in continents (c) and countries (d)

Thailand (451 km, nine users). However, there are less than 10 users in 11 out of those 17 countries, which explains the small total distance mapped for some of the countries.

Mapillary data can be contributed by local users or visitors, e.g. when vacationing. To identify a user's home country, we calculated a score based on a combination of days and distance mapped in each country. The marked peak for Australia/Oceania in November 2014 for distance per user (Figure 7c) is caused by a user with home country Austria who mapped actively for one week in Australia (Figure 8a). Another example of extensive visitor mapping can be found in the southwestern part of the Iberian Peninsula, where users from eight different countries, most of them probably tourists, contributed data (Figure 8b).

4.1.2 Individual level

Different measures can be considered to describe how individual mappers contribute data to Mapillary. Days of active contributions is a measure of a user's continued commitment to Mapillary data contribution. A related measure is the average distance mapped per week. It is calculated by dividing the total length of sequences uploaded by the number of weeks that passed since the first contribution of a user.

The spread of locations mapped by a user can provide some insight into travel behavior. One basic measure is the number of countries mapped by an individual user. A somewhat refined measure that uses coordinates of uploaded images and that was used in global mobility studies (González et al. 2008; Hawelka et al. 2014) is the radius of gyration. It measures how a user's data contributions are spread around the mass center as follows:

$$r=\frac{1}{n}\sqrt{\sum_{i=1}^{n}|p_i-\bar{p}|^2} \quad (3)$$

where n is the number of photo locations under consideration, p_i is the location of photo i, and $\bar{p}$ is the center of mass, calculated as the average of all photo locations.

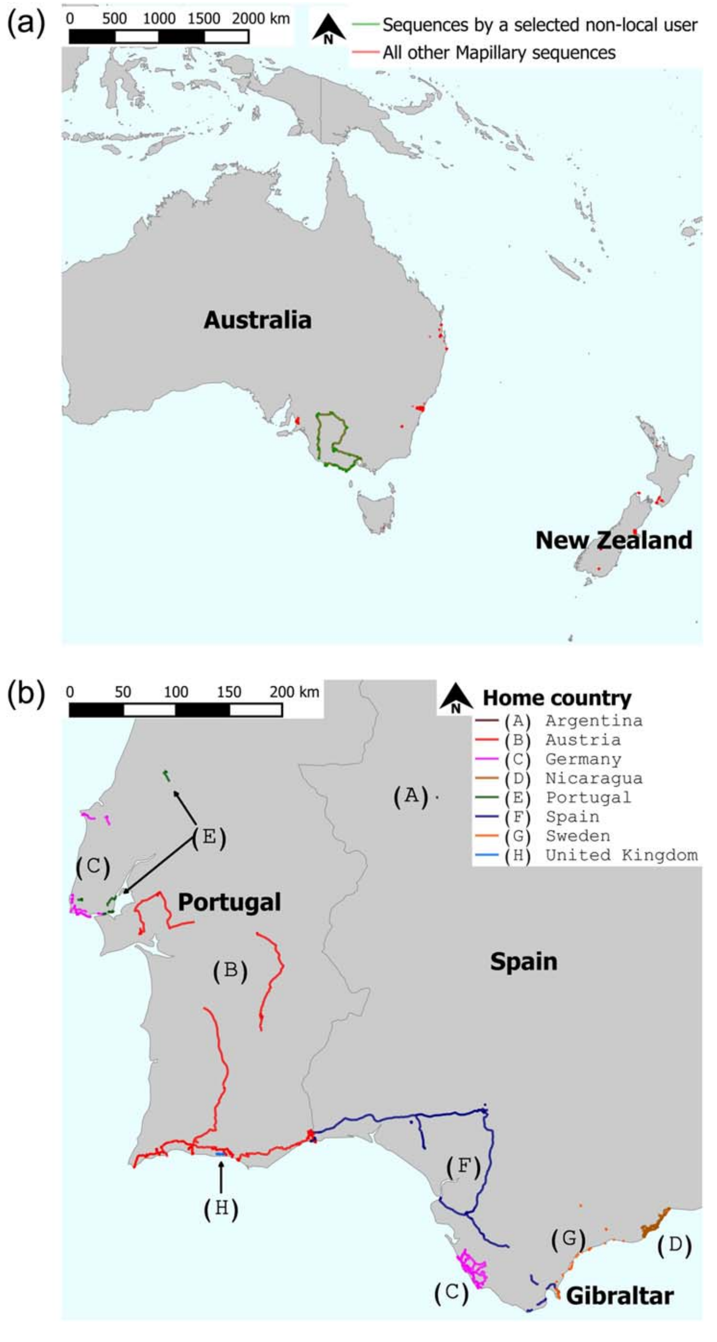

Figure 8 Sequences in Australia/Oceania (a) and the Iberian Peninsula (b)

For analyzing these variables, all users that registered at least 100 days before the end of the data collection (3 February, 2015) were selected. From this group, each user's contribution within the first 100 days since registration was analyzed. In this way, contribution patterns are not distorted by newly registered users who have not yet had a chance to contribute for as many days as other users. Average weekly distance was analyzed for the first 14 weeks of contribution, i.e. 98 days.

Low radius of gyration values indicate local contributions, whereas high values indicate long-distance travel for mapping purposes. The radius of gyration is not limited to contributions of individual users but can be aggregated for several users, e.g. for those residing in a specific country or continent. Analyzing those countries that were the home country to more than five users showed a mean value of 180 km (std. dev. = 241 km). In these countries, the radius of gyration was largest for Argentina (~610 km), followed by Ireland (~470 km), Austria (~360 km), Norway (~350 km), and Canada (~290 km). Whereas we hypothesized that country size, economic development, and island/mainland topology of a country would affect the size of the radius of gyration, multiple regression did not show any of these predictors to be significant.

The distributions of previously discussed contribution related variables can be well approximated by the power law function $P(\delta) = \delta \exp(-\beta)$, where δ is the contribution variable of interest and β is the exponent. Log-log plots in Figure 9 relate the proportion of users to the average weekly mapped distance (Figure 9a, $\beta = 0.96$), the number of active mapping days (Figure 9b, $\beta = 1.46$), the radius of gyration (Figure 9c, $\beta = 0.73$), and the proportion of users contributing on a specific day (first, second, etc.) since their initial contribution (Figure 9d, $\beta = 0.58$). Most contributors made their first uploads quickly after creating the account (median: 1.4 days). A power law relationship can be identified for the number of countries contributors mapped in ($\beta = 3.20$, R-squared: 0.91), where the maximum number of countries mapped per user was 10 (not shown in Figure 9). These patterns reveal participation inequality. For example, 1% of the users contributed data in over 50 days within their first 100 days of Mapillary, compared to an average of 6.1 days. Further, 1% of the selected users have a radius of gyration greater than 3,500 km, compared to an average radius of gyration of 130 km.

The exponent value for the radius of gyration was found to be somewhat larger when derived from Foursquare check-ins (Cheng et al. 2011) ($\beta = 0.99$) or georeferenced tweets (Hawelka et al. 2014) ($\beta = 1.25$). One possible explanation for this difference is that Mapillary contributors have to travel larger distances once their home region is mapped, whereas tweets and check-ins can be continuously posted from the same local surroundings.

Figure 9e shows which percentage of mappers contributes how long (up to 24 weeks) after their initial contribution. For this plot only users that had their first contribution at least 24 weeks before the end of the data collection were considered. The largest group contributes only up to three weeks after their initial contribution. The most committed group to the service (about 12%) is the one with a contribution span of 22-24 weeks and an average of 30 days of active mapping.

Figure 10 shows an overlay of convex hulls generated from each user's Mapillary data contributions. The map reveals main travel corridors between different geographic regions as part of the data collection process. The strongest ties can be identified between Europe and the US, where 31 mappers contributed data in both continents.

In the data collection efforts of individual mappers, two mapping strategies can be observed, closely resembling two mapper types previously identified by Heipke (2010). The first mapper type is a casual mapper, characterized as someone willing to spend only a little effort in mapping. Such a mapper would typically avoid extra trips just for the purpose of mapping but would be more likely to contribute photos along paths that are primarily traveled for some other reason, e.g. recreation along a hiking track (Figure 11a). The second type of mapper, the so called map lover, is someone who is used to working with correct maps throughout his/her professional life and who is more likely to collect data in a systematic

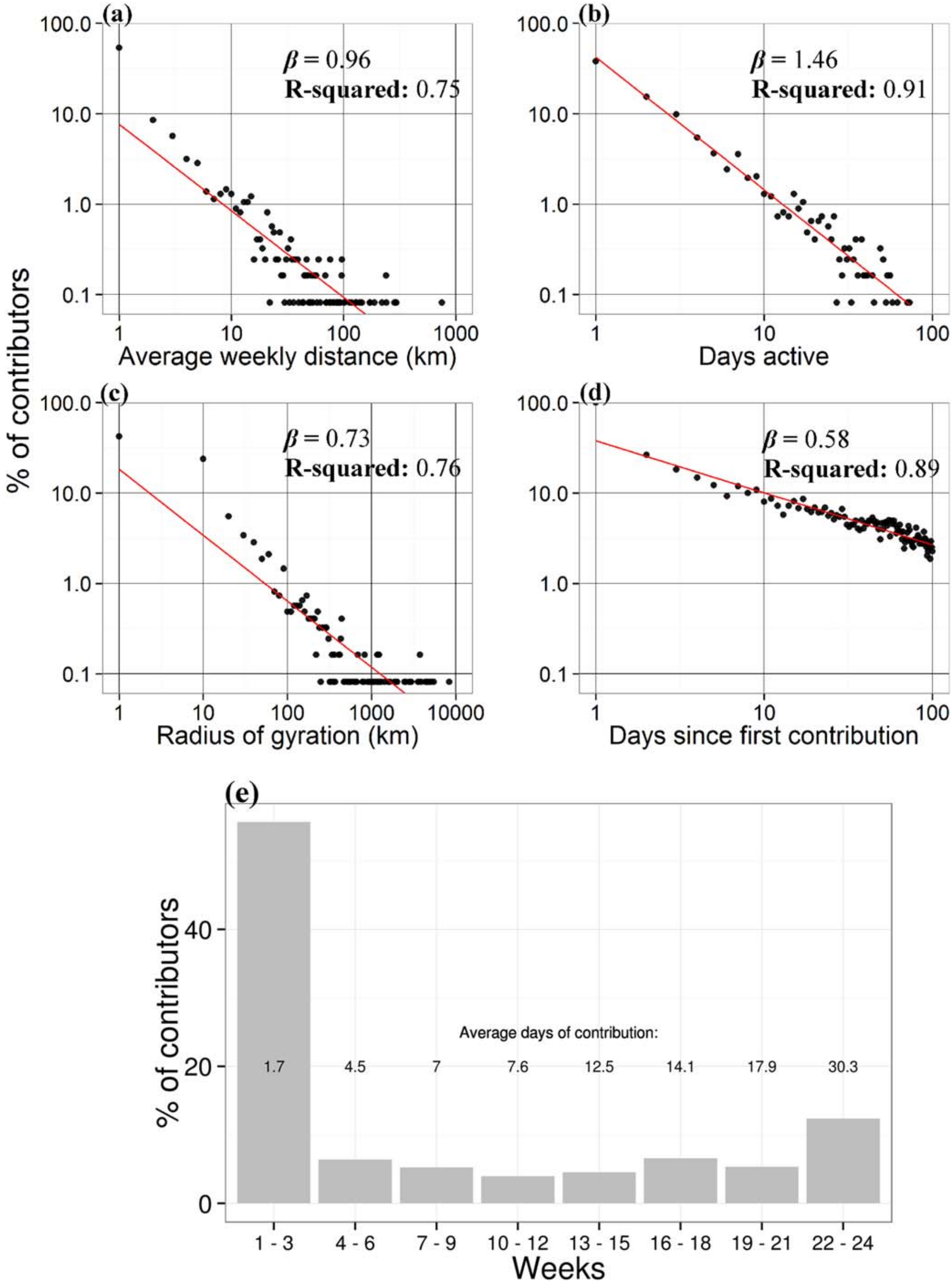

Figure 9 Power law approximations of contribution patterns for individual users (a-d) and contribution span of users (e)

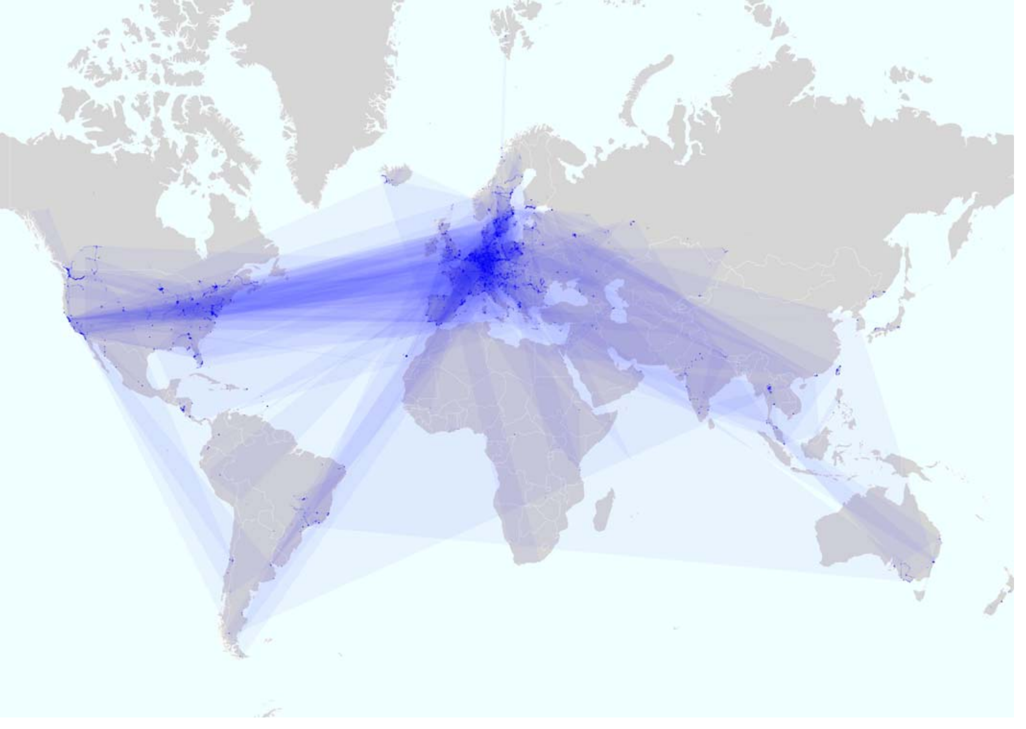

Figure 10 Convex hulls of contributions from individual users

manner. Figure 11b shows an example of a systematically mapped region where different colors indicate different Mapillary sequences uploaded by the same user.

Besides contributing new data users can also review photos and sequences uploaded by others as part of the quality improvement process. This includes correcting orientation angles, editing blurs and street signs, and hiding private photos or photos of poor quality. These changes need to be manually approved by Mapillary. There is even a group of users who focus more on data edits than on photo uploads. Although data edits are permitted, Mapillary does not provide a public editing history of images at this point. As opposed to this, in other VGI sources, such as OSM, the object editing history is frequently used for quality assessment of mapped features (Neis and Zielstra 2014; Rehrl et al. 2013).

4.2 Completeness of Street Level Photos

4.2.1 Completeness of Mapillary data at the country level

Completeness is computed by comparing the amount of mapped road data in Mapillary to a reference dataset. Figure 12 shows the completeness of Mapillary on main roads for each country, where Esri main roads were used as the reference dataset. Sixteen out of the 20 most completely mapped countries are in Europe. However, the highest completeness values can be found for three individual administrative units outside Europe (Table 1). These are Barbados, Hong Kong, and Nicaragua. Whereas Barbados is a small country where mapping 11 km of main roads already leads to a completeness value of over 50% (first row), the high

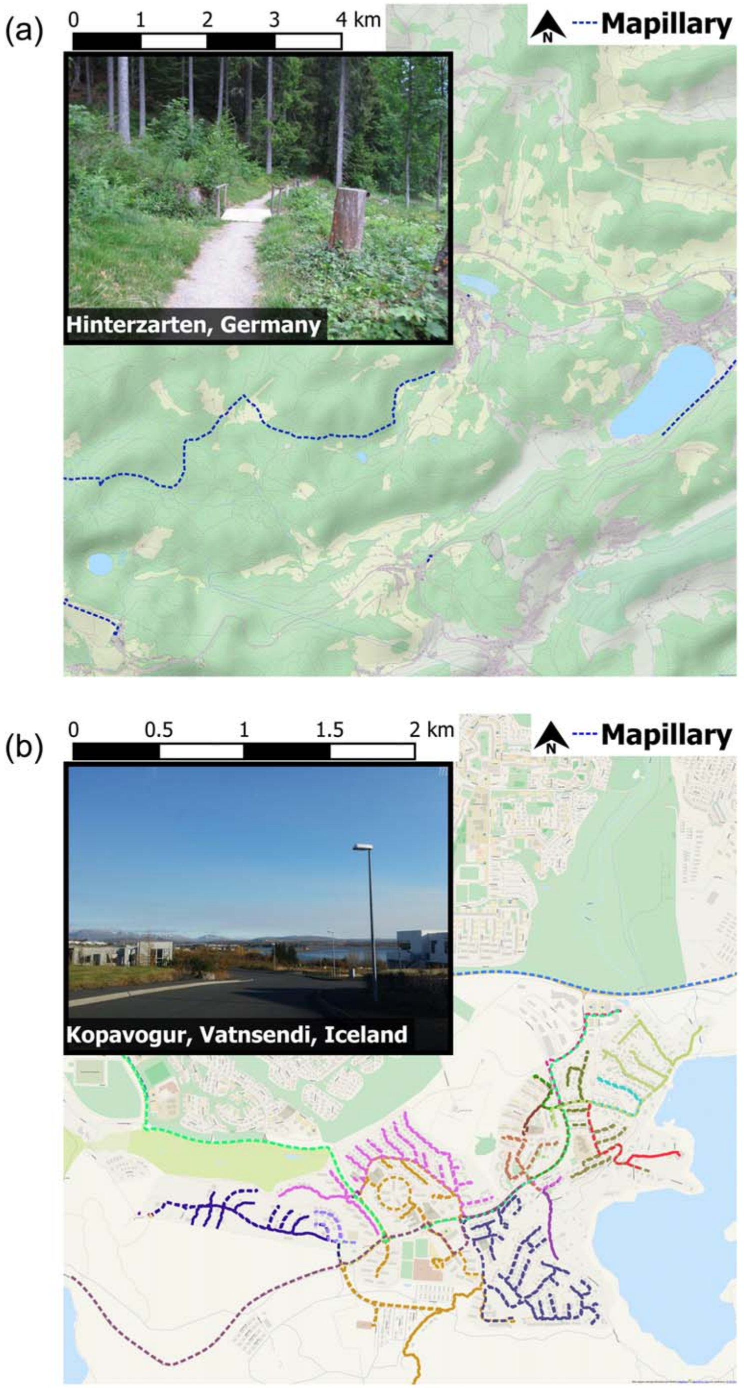

Figure 11 Contributions of different mapper types: Casual mapper recording a single hiking trail (a) and map lover systematically recording a neighborhood (b)

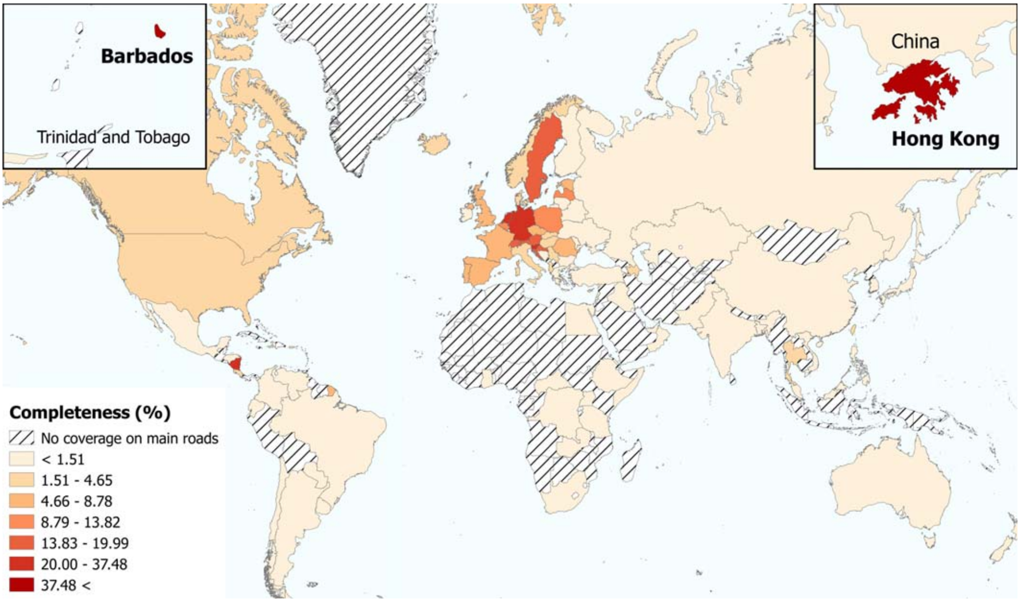

Figure 12 Completeness of Mapillary on main roads

Table 1 Most complete administrative units

Administrative unit	Length of main road system [km]	Mapped main road distance [km]	Completeness [%]	Number of users
Barbados	21	11	52.6	2
Hong Kong	213	110	51.5	2
Nicaragua	1,365	512	37.5	6
Slovenia	1,449	456	31.5	9
Netherlands	4,770	1,426	29.9	49
Germany	42,893	11,452	26.7	351
Croatia	4,363	874	20.0	15
Sweden	18,124	3,534	19.5	301
Austria	5,382	1,001	18.6	49
Switzerland	3,554	608	17.1	49

completeness rates in Hong Kong and Nicaragua can be clearly attributed to power users, where two mappers contributed 110 km (Hong Kong), and six mappers contributed 512 km (Nicaragua), respectively, which results in a higher average contribution rate per user than for the country with the largest total mapped main road distance, which is Germany.

4.2.2 Completeness of Mapillary and Street View data at the city level

For 11 selected urban, mixed, and rural areas in the US and northern Europe we calculated the completeness of Mapillary and Street View as well as the relative completeness difference between Mapillary and Street View. All measures were computed separately for OSM main roads, residential roads, and off-road pedestrian/bicycle features (Table 2). At the city level, Street View provides more complete imagery for all study areas on all road categories, except for off-road features in Malmö, Sweden. For the selected areas Street View imagery provides, on average, high coverage of main roads (95.8%), followed by fairly complete coverage of residential roads (78.5%) and much lower coverage of off-road segments (7.6%). The two study sites with rural characteristics show for all road types much lower completeness values than the mixed and urban neighborhoods, where completeness values of mixed areas lie between those of urban and rural neighborhoods. Hence the 11 analyzed stud regions suggest a decline in Street View coverage from urban to rural areas.

In Mapillary, similar completeness patterns can be observed. That is, main roads are most completely mapped (28.0%), followed by residential roads (3.8%) and off-road segments (2.3%). Completeness ranges between 10.1% (Los Angeles) and 58.0% (Malmö) for main roads, between 1.3% (Fort Lauderdale, Minneapolis) and 20.4% (Malmö) for residential roads, and between 0.0% (Fort Lauderdale) and 10.3% (Malmö) for off-road segments. Malmö, as the home town of the Mapillary project, provides significantly better coverage than other cities. It more than doubles the completeness value for residential roads and off-road segments of the second best mapped city in the corresponding categories. This data abundance results in relative completeness difference values that are smaller than for other cities, meaning that Street View is less dominant than Mapillary in this city than in others. The relative completeness difference for off-road pedestrian/cycling features is even negative with a value of -0.14, showing that in Malmö more pedestrian/cycling features are mapped in Mapillary than in Street View.

Figure 13 visualizes for selected areas in Table 2 patterns of relative completeness differences for different road categories. Purple tiles without borders show areas with more roads mapped in Street View, whereas orange tiles with borders indicate areas with higher Mapillary coverage. For rural areas in Washington state (Figure 13a, relative completeness difference on main roads: 0.22), most of the main roads were covered by both imagery (white tiles). Street View imagery is available on more main roads than Mapillary (purple tiles), but some main roads are only covered in Mapillary (orange tiles). In the second rural study area, for residential roads (Figure 13b, relative completeness difference on residential roads: 0.85) several road segments in the center are covered only by Mapillary but not by Street View. Figure 13c shows a case where a single spot provides better Mapillary than Street View coverage. The relative completeness difference for the Vatnsendi neighborhood in the southeast of Reykjavik is -0.1 as opposed to the observed value of 0.91 for the whole city. This is a result of the previously identified map lover user who systematically mapped the whole neighborhood in October 2014 (compare also Figure 11b).

Numerous orange tiles in Figure 13d for Malmö demonstrate the better coverage of Mapillary over Street View for off-road segments in the city as a whole. Figure 13e illustrates for San Francisco that even in urban areas where Street View captures a high number of off-road features, Mapillary can provide better coverage in selected districts.

5 Conclusions and Future Work

This article analyzed how users have contributed street level photographs to Mapillary throughout the first year since the inception of this project. Assessment of data quality is especially crucial for VGI data sources since these data are not subject to official quality standards posed by regulatory agencies. Mapillary data, like most other VGI data, do not come with traditional measures of accuracy in their metadata. The article assessed the completeness of

Table 2 Completeness of Mapillary (Map) and Street View (SV) together with relative completeness difference

Type	Study sites	Area [km^2]	OSM main			OSM residential			OSM off-road		
			SV	Map	Rel. Diff.	SV	Map	Rel. Diff.	SV	Map	Rel.Diff.
Urban	Los Angeles, CA	885	99.3	10.1	0.81	93.5	1.6	0.97	7.3	4.1	0.27
	Ft. Lauderdale, FL	279	98.9	24.1	0.61	88.4	1.3	0.97	8.3	0.0	1.0
	Downtown Miami, FL	97	99.6	40.0	0.42	91.6	6.8	0.86	7.0	2.6	0.45
	New York City, NY	190	99.0	19.1	0.68	95.5	6.6	0.87	40.4	2.4	0.89
	San Francisco, CA	447	99.5	37.1	0.46	92.6	4.8	0.90	15.0	1.9	0.78
		Average	**99.2**	**21.8**		**93.0**	**3.2**		**15.9**	**3.2**	
Rural	Bellingham, WA area	2,569	75.3	47.9	0.22	26.2	6.9	0.58	4.0	0.6	0.75
	Eldersburg, Maryland	2,451	87.4	41.0	0.36	53.7	4.2	0.85	1.0	0.5	0.34
		Average	**83.2**	**43.7**		**41.8**	**5.6**		**2.1**	**0.5**	
Mixed	Malmö, Sweden	409	97.1	58.0	0.25	75.0	20.4	0.57	7.7	10.3	−0.14
	Minneapolis, MN area	4,126	97.6	20.8	0.65	81.6	1.3	0.97	11.5	0.6	0.90
	Reykjavik, Iceland	905	96.1	25.3	0.58	82.1	3.8	0.91	5.4	0.7	0.76
	Washington, DC area	912	98.9	26.7	0.58	89.8	4.0	0.92	7.7	1.4	0.69
		Average	**98.0**	**25.8**		**84.6**	**3.4**		**7.3**	**2.6**	
	Total average		**95.8**	**28.0**		**78.5**	**3.8**		**7.6**	**2.3**	

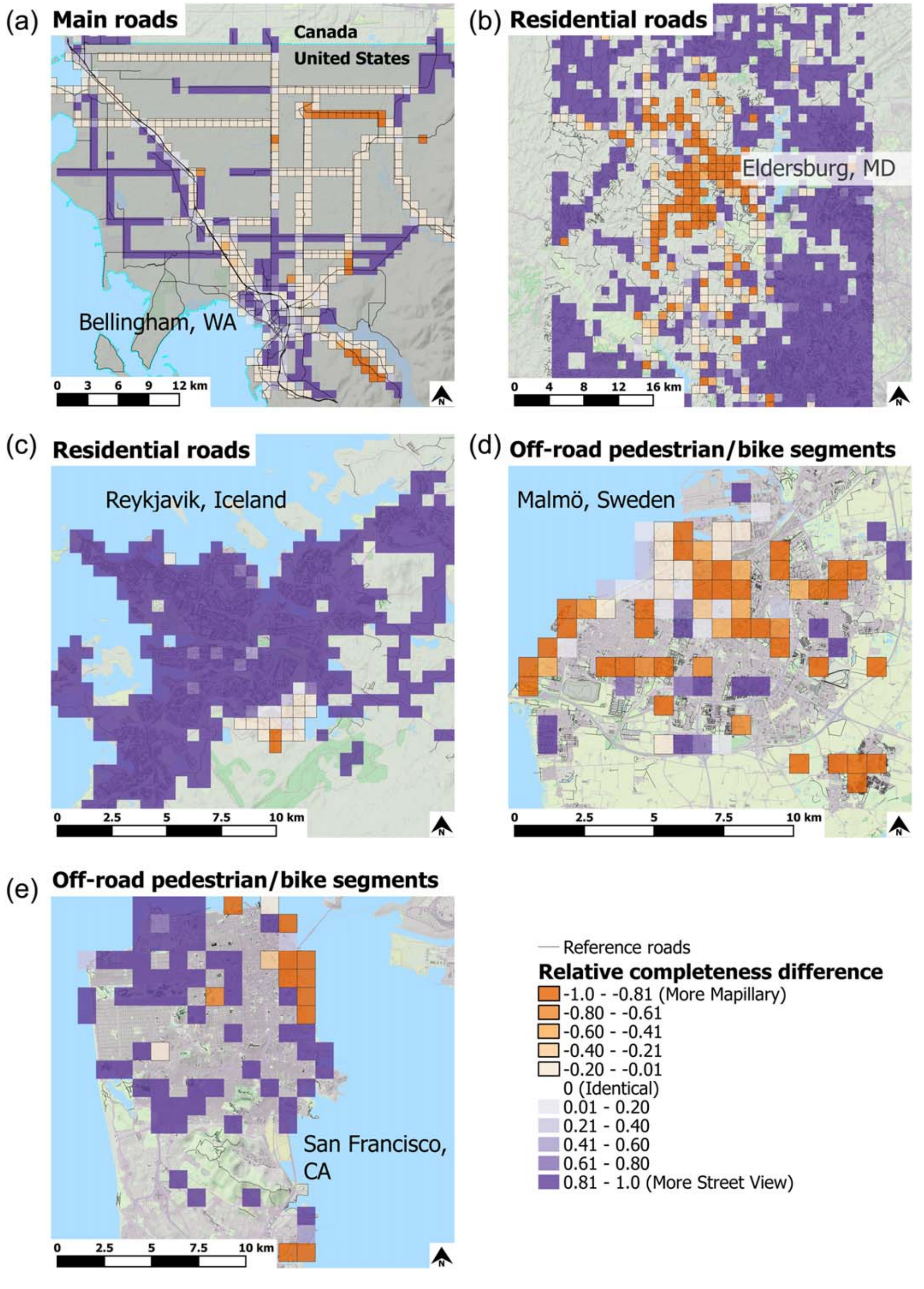

Figure 13 Spatial distribution of relative completeness difference for selected cities

Mapillary on main roads for countries from all over the world, and compared for selected study regions completeness values of Mapillary to those of Street View for three road categories. For all 11 analyzed local areas, Street View delivers better coverage than Mapillary with one exception for Malmö in combination with off-road features. However, both Street View and especially Mapillary imagery is not evenly distributed within study sites, so that regional pockets exist where Mapillary coverage outperforms that of Street View in the various analyzed road categories. These pockets could indicate either a heightened interest from users in mapping a specific area, or a gap in Street View coverage. The presented analysis method, however, does not allow to distinguish between these two cases.

The Mapillary service experiences a high data growth rate. Although one cannot predict future growth rates, the analyses identified a strongly committed group of users that contribute data on a regular basis after joining the Mapillary project. The distribution of contribution related measures (days of active mapping, mapped kilometers per week, etc.) follow closely the power law function, reflecting distribution inequality. This kind of participation inequality matches contribution patterns observed for other VGI data sources, and can be explained by different motivations for different kinds of contributors. For example, Budhathoki and Haythornthwaite (2013) distinguish between lightweight (casual) and heavyweight (serious) mappers in OSM. More specifically, a casual mapper's motivation is more oriented towards general principles of free availability of mapping data, whereas serious mappers are primarily motivated by gaining knowledge and advancing their career. Analysis of the radius of gyration reveals that Mapillary is not limited to local activities, but that some users travel and map locations far apart. In fact, 18% of all contributors have a radius of gyration greater than 100 km.

Availability of Mapillary imagery through the ShareAlike license can facilitate various research efforts that require such imagery (e.g. wheelchair routing), as well as the development of location based services (e.g. virtual tours of hiking trails). Extraction of traffic signs through computer vision algorithms can become helpful for managing transportation utility inventory. Despite these applications, due to the crowdsourced nature of Mapillary, usability for this kind of purpose is limited to focus areas only.

For future analysis, we plan to analyze how OSM contributors use Mapillary imagery as a data source. Mapillary photos are already included in the official OSM editor so that OSM mappers have access to street level photos. Tags in the source field of OSM features indicate that OSM mappers already use Mapillary imagery to add map features, such as bus stops, which can be visually identified on Mapillary images. It will also be of interest to analyze whether the same group of volunteer mappers contributes to OSM and Mapillary or if a new crowd of voluntary mappers are reached via Mapillary.

Further we plan to refine the categorization of volunteer mappers and extend user types beyond those who actively contribute photos. It can, for example, be observed that a group of users provide major contributions without uploading images, but rather through editing of existing images. These edits include blurring requests for faces and license plates, correction of camera angles, or splitting long sequences to match road segments rather than representing a whole trajectory.

References

Backstrom L, Sun E, and Marlow C 2010 Find me if you can: Improving geographical prediction with social and spatial proximity. In Proceedings of the Nineteenth International Conference on World Wide Web, Raleigh, North Carolina: 61–70

Budhathoki N R and Haythornthwaite C 2013 Motivation for open collaboration crowd and community models and the case of OpenStreetMap. *American Behavioral Scientist* 57: 548–75

Cheng Z, Caverlee J, Lee K, and Sui D Z 2011 Exploring millions of footprints in location sharing services. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain: 81–8

Cipeluch B, Jacob R, Winstanley A, and Mooney P 2010 Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. In *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Leicester, UK: 337–40

Clarke P, Ailshire J, Melendez R, Bader M, and Morenoff J 2010 Using Google Earth to conduct a neighborhood audit: Reliability of a virtual audit instrument. *Health and Place* 16: 1224–29

Davis Jr C A, Pappa G L, de Oliveira D R R, and de L Arcanjo F 2011 Inferring the location of twitter messages based on user relationships. *Transactions in GIS* 15: 735–51

Elwood S, Goodchild M F and Sui D Z 2012 Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the Association of American Geographers* 102: 571–90

Girres J F and Touya G 2010 Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS* 14: 435–59

Gonzalez A, Bergasa L M, and Yebes J J 2014 Text detection and recognition on traffic panels from street-level imagery using visual appearance. *IEEE Transactions on Intelligent Transportation Systems* 15: 228–38

González M C, Hidalgo C A, and Barabási A-L 2008 Understanding individual human mobility patterns. *Nature* 453: 779–82

Goodchild M F 2007 Citizens as voluntary sensors: Spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research* 2: 24–32

Google 2015 Maps JavaScript API v3. WWW document, http://developers.google.com/maps/documentation/javascript/maptypes#TileCoordinates

GRASS Development Team 2015 GRASS GIS 7.1 Reference Manual. WWW document, http://grass.osgeo.org/grass71/manuals/

Griew P, Hillsdon M, Foster C, Coombes E, Jones A, and Wilkinson P 2013 Developing and testing a street audit tool using Google Street View to measure environmental supportiveness for physical activity. *International Journal of Behavioral Nutrition and Physical Activity* 10(1): 103

Guy R and Truong K 2012 CrossingGuard: Exploring information content in navigation aids for visually impaired pedestrians. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Austin, Texas: 405–14

Haklay M 2010 How good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B* 37: 682–703

Hara K, Sun J, Moore R, Jacobs D, and Froehlich J 2014 Tohme: Detecting curb ramps in Google Street View using crowdsourcing, computer vision, and machine learning. In *Proceedings of the Twenty-seventh Annual ACM Symposium on User interface Software and Technology*, Honolulu, Hawaii: 189–204

Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, and Ratti C 2014 Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science* 41: 260–71

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*, Savannah, Georgia: 229–32

Heipke C 2010 Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 550–57

Hochmair H H and Zielstra D 2015 Analysing user contribution patterns of drone pictures to the Dronestagram photo sharing portal. *Journal of Spatial Science* 60: 79–98

Hochmair H H, Zielstra D, and Neis P 2015 Assessing the completeness of bicycle trail and lane features in OpenStreetMap for the United States. *Transactions in GIS* 19: 63–81

Javanmardi S, Ganjisaffar Y, Lopes C, and Baldi P 2009 User contribution and trust in wikipedia. In *Proceedings of the Fifth International IEEE Conference on Collaborative Computing: Networking, Applications and Worksharing*, Washington, DC: 1–6

Juhász L and Hochmair H H 2015 Exploratory completeness analysis of Mapillary for selected cities in Germany and Austria. In Car A, Jekel T, Strobl J, and Griesebner G (eds) *GI_Forum 2015 – Geospatial Minds for Society*. Berlin, Wichmann: 535–45

Li L, Goodchild M F, and Xu B 2013 Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science* 40: 61–77

Masó J, Pomakis K, and Julià N 2010 OpenGIS Web Map Tile Service Implementation Standard v1.0.0. Wayland, MA, Open Geospatial Consortium Technical Document No. 07-057r7

Neis P and Zielstra D 2014 Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet* 6: 76–106

Neis P, Zielstra D, and Zipf A 2013 Comparison of volunteered geographic information data contributions and community development for selected world regions. *Future Internet* 5: 282–300

Neis P and Zipf A 2012 Analyzing the contributor activity of a volunteered geographic information project: The case of OpenStreetMap. *ISPRS International Journal of Geo-Information* 1: 146–65

Nielsen J 2006 The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities. WWW document, http://www.nngroup.com/articles/participation-inequality/

OSM 2015 OSM Wiki. WWW document, http://wiki.openstreetmap.org/wiki/Slippy_map_tilenames#X_and_Y

Rehrl K, Gröchenig S, Hochmair H H, Leitinger S, Steinmann R, and Wagner A 2013 A conceptual model for analyzing contribution patterns in the context of VGI. In Krisp J M (ed) *Progress in Location-Based Services*. Berlin, Springer: 373–88

Rundle A G, Bader M D M, Richards C A, Neckerman K M, and Teitler J O 2011 Using Google Street View to audit neighborhood environments. *American Journal of Preventive Medicine* 40: 94–100

Vanwolleghem G, Dyck D V, Ducheyne F, Bourdeaudhuij I D, and Cardon G 2014 Assessing the environmental characteristics of cycling routes to school: A study on the reliability and validity of a Google Street View-based audit. *International Journal of Health Geographics* 13: 19

Yin L, Cheng Q, Wang Z, and Shao Z 2015 'Big data' for pedestrian volume: Exploring the use of Google Street View images for pedestrian counts. *Applied Geography* 63: 337–45

Zamir A R and Shah M 2010 Accurate image localization based on Google Maps Street View. In Daniilidis K, Karagos P, and Paragios N (eds) *Computer Vision – ECCV 2010*. Berlin, Springer: 255–68

Zielstra D and Hochmair H H 2011 A comparative study of pedestrian accessibility to transit stations using free and proprietary network data. *Transportation Research Record* 2217: 145–52

Zielstra D, Hochmair H H, Neis P, and Tonini F 2014 Areal delineation of home regions from contribution and editing patterns in OpenStreetMap. *ISPRS International Journal of Geo-Information* 3: 1211–33

Zielstra D and Zipf A 2010 OpenStreetMap data quality research in Germany. In *Proceedings of the Sixth International Conference on Geographic Information Science*, Zurich, Switzerland